Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

http://csli-publications.stanford.edu/LFG/2019

Contents

1 Editor's Note

The 2019 Conference on Lexical Functional Grammar was held at the Australian Natural University in Canberra, Australia. The program committe for LFG19 were John Lowe and Agnieszka Patejuk. We would like to thank them for coordinating the review process and working with the local and workshop organizers to put together this year's broad program, which also included a teach-in. We would also like to thank the executive committee and the abstract and final paper reviewers, without whom the conference and the proceedings would not have been possible in this form.

The local organization was taken care of by the dedicated efforts of I Wayan Arka, Elisabeth Mayer, Jane Simpson, Avery Andrews and Cynthia Allen, whom we would like to thank for a wonderfully organized conference, rounded off with a bushwalk full of kangaroos and a wonderful dinner at the planetarium. Our thanks go out to them for their very hard work in ensuring a successful, varied and enjoyable conference.

This year's conference included a Workshop on the morphology-syntax interface in underdescribed languages and a teach-in on historical linguistics from an LFG perspective. Submitted papers from these events have been included in the proceedings.

The table of contents lists all the papers presented at the conference. Some papers were not submitted to the proceedings. For these papers, we suggest contacting the authors directly. We note that all of the abstracts were peer-reviewed anonymously (double-blind reviewing) and that all of the papers submitted to the proceedings underwent an additional round of reviewing. We would like express our heartfelt thanks to all of the anonymous reviewers for the donation of their expertise and effort in what is often a very short turn-around time.

Hard Copy: All of the papers submitted to the LFG19 proceedings are available in one large pdf file, to be viewed and printed with Adobe Acrobat. The proceedings' file was created via pdflatex tools and a script written by Stefan Müller. We are highly indebted to him for the use of the script. We thank Sarah Weaver at CSLI Publications for making sure the proceedings become accessible via the CSLI site. Finally, we thank Dikran Karagueuzian at CSLI Publications for his continuous support of our proceedings and our community.

Part I
Contributions to the Main Conference

Constraining Expletives in English

Alex Alsina
Universitat Pompeu Fabra

Fengrong Yang
Universitat Pompeu Fabra

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 6–26

http://csli-publications.stanford.edu/LFG/2019

Keywords: English, expletives, mapping theory

Alsina, Alex, & Yang, Fengrong. 2019. Constraining Expletives in English. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 6–26. Stanford, CA: CSLI Publications.

Abstract

This paper aims to present a theory of expletives[1] in English within LFG in which the distribution of expletives follows from general principles and from the lexical entries of the relevant expletives. Consequently, expletives are not subcategorized for, i.e. verbs do not lexically specify whether they take an expletive or what expletive they take, unlike what is assumed in current LFG approaches to expletives. In addition, there are no alternative lexical entries for verbs depending on whether they cooccur with an expletive or not. The proposed analysis diverges from standard mapping theories in the assumption that argument-to-function linking takes place in the syntax, instead of in the lexicon. The current analysis assumes that there are two kinds of expletives: non-thematic expletives, which do not fill an argument position and are licensed by satisfying the Subject Condition, and argumental expletives, which do fill an argument position, but make no semantic contribution.

1 Introduction

The goal of this paper is to propose an alternative theory to the current accounts of the English expletives *there* and *it*, a theory that does not account for the distribution of these expletives through the stipulation of the specific expletive required. This paper is organized as follows. In section 2, we review current analyses of English expletives *there* and *it* and point out the problems they face. We propose our theory about expletive *there* and propositional *it* in section 3 and 4, respectively. Main conclusions are summarized in section 5.

2 Current analyses of expletives

A commonly accepted assumption in LFG is that an expletive GF has to be included in the lexical entry of the predicate that cooccurs with this GF in the clause. For example, Bresnan (1982), in addition to positing the lexical entry of expletive *there*, as in (1a), assumes that the lexical entry of copular *be*, as used in *There is a pig running through the garden*, includes only one thematic argument in its PRED value (i.e., the XCOMP), and has to stipulate that **i)** it takes a non-thematic SUBJ, **ii)** this SUBJ is an expletive, and **iii)** the form of the expletive is *there*, represented as in (1b):

(1) a. *there*: $\text{NP}_{[\text{PRO}]}$, (FORM) = there
    b. V, 'there-be ((XCOMP))', X = P, V, A
        (OBJ) = (XCOMP SUBJ)
        (SUBJ FORM) $=_C$ there
        (SUBJ NUM) = (OBJ NUM)

(Bresnan 1982: 73)

† We deeply acknowledge the comments and observations made by anonymous reviewers and the audience of the 24th LFG conference. We especially thank Joan Bresnan, Mary Dalrymple, and Péter Szücs for their useful suggestions. Any remaining errors are our own.

[1] By an expletive we refer to a grammatical function (GF) with no semantic content. Weather "it" will not be considered as an instance of an expletive, given compelling arguments for its non-expletive status (see for example Krejci 2014, Levin 2017, etc.).

Falk (2001) takes a similar position, proposing that in constructions in which an expletive subject cooccurs with a clausal complement, the main predicate should be analyzed as subcategorizing for a COMP and an expletive subject, with a special requirement on the form of the expletive. Namely, the non-thematic subject must be realized as *it* in such constructions:

(2) It seems that the geneticist clones dinosaurs.

(Falk 2001: 137)

The lexical entry of *seem* in (2), with the requirement on the form of the non-thematic expletive subject, is suggested to be represented as follows:

(3) *seem*: V (↑ PRED) = 'seem <(↑ COMP)> (↑ SUBJ)'
(↑ SUBJ FORM) = it

(Falk 2001: 138)

As an alternative to stipulating the specific expletive required by means of the feature FORM, which is also present in the lexical entry of the expletive (as in (1a)), some LFG theories achieve the same result by stipulating in the lexical entry of a predicate the person, number, and gender features of one of the GFs required by the predicate, which can only match those of a particular expletive. For example, Kaplan and Zaenen (1995) propose that the predicate *likely* as used in example (4a) has the lexical entry in (4b) and the expletive *it* has the lexical entry in (4c).

(4) a. It is likely that Susan will be late.
b. *likely*: (↑ PRED) = 'likely <(↑ COMP)> (↑ SUBJ)'
(↑ SUBJ PERS) = 3
(↑ SUBJ NUM) = SG
(↑ SUBJ GEND) = NEUT
c. *it*: (↑ PERS) = 3
(↑ NUM) = SG
(↑ GEND) = NEUT
¬ (↑ PRED)

(Kaplan and Zaenen 1995: 158)

A consequence of stipulating the presence of an expletive in the lexical entry of a verb is that the distribution of expletives does not follow from general principles (such as the Subject Condition), unlike what happens in GB/MP, where it is a direct consequence of the Extended Projection Principle (EPP, Chomsky 1981, 1982, etc.). Moreover, it does not provide a way of explaining the distribution of the expletives *there* and *it*, namely, the fact that, in standard modern English, expletive *there* cooccurs with a postverbal NP, whereas expletive *it* cooccurs with phrases with propositional content, such as infinitival phrases or full clauses:[2]

(5) a. There/*it are flowers in the yard.

---

[2] Expletive *it* was also used in existential constructions in earlier stages of English, and is still used in African American Vernacular English (Louise McNally, p.c.).

b. It/*there seems that a new idea is emerging.
c. It/*there surprised me that you won the lottery.

Another implication of the assumption in current LFG that expletives are lexically selected is that there are two lexical entries for predicates that may take an expletive as their subject. In other words, predicates must have two different lexical entries depending on whether they use an expletive or not. This is the case for Kaplan and Zaenen (1995) with “extraposable” predicates such as *likely*, *important*, and *advisable*, etc., which are proposed to have two different lexical entries, one with and one without the expletive subject *it*:

(6) a. That Susan will be late is likely.
likely: (↑ PRED) = ‘likely <(↑ SUBJ)>’
b. It is likely that Susan will be late.
likely: (↑ PRED) = ‘likely <(↑ COMP)> (↑ SUBJ)’

Generation of the lexical entry in (6b) is achieved by positing an extraposition rule for “extraposable” lexical entries, which licenses a second lexical entry with a non-thematic subject (for another proposal within LFG, see Falk 2001):

(7) Extraposition rule:
a. Extraposable entry:
(↑ PRED) = ‘R <(↑ SUBJ) … >’
b. Lexical entry added:
(↑ PRED) = ‘R <(↑ COMP) … > (↑ SUBJ)’
(↑ SUBJ PERS) = 3
(↑ SUBJ NUM) = SG
(↑ SUBJ GEND) = NEUT
(Kaplan and Zaenen 1995: 158)

However, the extraposition rule proposed by Kaplan and Zaenen (1995) can only account for constructions in which the original *subject* clause is extraposed. Such a rule gives no explanation, for example, for constructions in which an original object clause is extraposed, as in the following examples:

(8) a. I resent it greatly that you didn’t call me.
b. I regret it very much that we could not hire Mosconi.
(Postal and Pullum 1988: 642)

In summary, current LFG accounts of expletive GFs depend largely on stipulations in the lexical entries of the expletive-taking predicates. In addition, attempts to capture generalizations by means of lexical rules are partial (as they only address extraposition *it*) and incomplete (as they do not generalize to all instances of extraposition *it*).

3 Analysis of expletive *there*

In this section, we propose our analysis of expletive *there*. We assume that the distribution of expletive *there* is regulated by its lexical entry and by independently required constraints, such as the Subject Condition. For instance, verbs such as *appear* or *stand* allow their NP argument to be realized alternatively as SUBJ or OBJ, and when the OBJ realization is chosen, the expletive *there* is the syntactic SUBJ. This is illustrated by examples in (9) and (10), respectively:

(9) a. A bird appeared on the windowsill.
  b. There appeared a bird on the windowsill.

(10) a. A monument stood in the square.
  b. There stood a monument in the square.

We propose that verbs like *appear* or *stand,* despite having two alternative realizations of their NP argument, have only one lexical entry, with an a-structure consisting of one core argument, as shown in (11) for *appear*.

(11) ‘*appear* $<$ I A $>$’[3]

The mapping principles allow for an internal argument to map onto either SUBJ or OBJ, as assumed in current versions of mapping theories such as LMT (see Bresnan 1994, Kibort 2001, or Findlay 2016) or FMT (see Alsina 1996a). Therefore, the a-structure in (11) is used in all the c-structure/f-structure pairs corresponding to the examples in (9). As for the association between arguments and GFs, we are in line with FMT in assuming it takes place in the f-structure and not in the lexicon.

Let us assume the following two well-formedness conditions on f-structures, as violable constraints, i.e., the Subject Condition (SUBJCON) and GF Faithfulness (GF-FAITH), stated respectively as:

(12) Subject Condition (SUBJCON):
Every verbal f-structure must include a subject.

(13) GF Faithfulness (GF-FAITH):
i) Every direct GF must be lexically required (i.e., required by the lexical information of some element of the clause, such as the a-structure of the predicate).
ii) A GF has a PRED value iff it corresponds to an argument or has semantic content.

GF-FAITH is roughly equivalent to the commonly assumed Coherence Condition (Kaplan and Bresnan 1982) and replaces it in the present theory. Adopting an Optimality Theory (OT) approach to constraint interaction (see Kuhn 2003), two rankings of these constraints are possible, listed as follows:

(14) a. Ranking 1: SUBJCON $\gg$ GF-FAITH
  b. Ranking 2: GF-FAITH $\gg$ SUBJCON

With Ranking 1, we have languages in which every clause must include a subject, even if that incurs a violation of GF-FAITH, as in the case of a structure with an expletive subject that is not lexically required. English is an example of such a language. By contrast, with Ranking 2, we have languages in which expletive subjects are not possible, as every direct GF must be

[3] Here we use *I* to represent the internal argument and *A* to represent the non-core locative argument. *E* will be used to represent the external argument in (24).

lexically required (typically, an argument in the a-structure), even if that implies violating SUBJCON. Spanish or Catalan is an example of such a language.

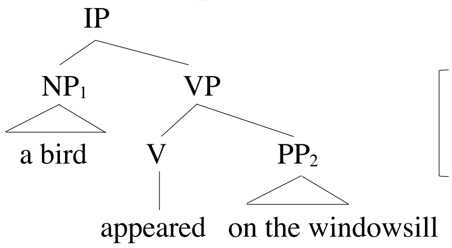
As for the lexical entry of expletive *there*,[4] we assume that it merely states that its category is NP and that it maps onto a GF with oblique case and that it cooccurs with an NP object in the same clause.[5]

(15) *there*: NP_1 $\begin{bmatrix} \mathrm{GF} \begin{bmatrix} \mathrm{CASE} & \mathrm{OBL} \end{bmatrix}_1 \\ \mathrm{OBJ} \end{bmatrix}$

where OBJ maps onto an NP

Let us consider the representation of the examples in (9). If we choose to map the internal argument of the predicate in (11) to SUBJ, we obtain the c-structure/f-structure pair in (16), corresponding to example (9a). Here and in what follows, the subscripted integers signal the correspondence between arguments, GFs, and c-structure nodes:

(16) C-structure/f-structure pair with the internal argument (I) as SUBJ:



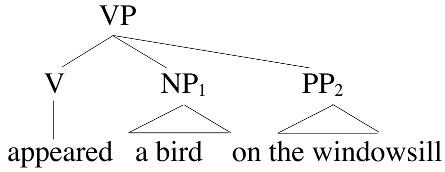
$$\begin{bmatrix} \mathrm{PRED} & \text{'appear} < \mathrm{I}_1 \;\; \mathrm{A}_2 > \text{'} \\ \mathrm{SUBJ}_1 & \\ \mathrm{OBL}_2 & \end{bmatrix}$$

This is the optimal candidate: the f-structure in (16) satisfies both SUBJCON and GF-FAITH, because it contains a subject which is lexically required (it maps onto an argument).

If we choose the option in which the internal argument of the predicate in (11) maps to OBJ, there are two possible f-structures consistent with this mapping, as shown in the two c-structure/f-structure pairs in (17):

(17) Possible c-structure/f-structure pairs with internal argument (I) as OBJ:

a.

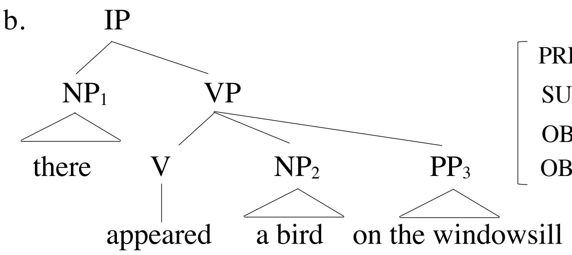


$$\begin{bmatrix} \mathrm{PRED} & \text{'appear} < \mathrm{I}_1 \;\; \mathrm{A}_2 > \text{'} \\ \mathrm{OBJ}_1 & \\ \mathrm{OBL}_2 & \end{bmatrix}$$

[4] We are not using annotations for lexical entries, but we give the f-structures that would be generated by the annotations.

[5] The reference to oblique case in the lexical entry of expletive *there* serves to block verb agreement with the expletive subject, allowing the direct case object to be the agreement trigger, as in Bresnan (1994).

b.



$$\begin{bmatrix} \text{PRED} & \text{'appear} < \text{I}_1 \ \text{A}_2 > \text{'} \\ \text{SUBJ} & [\text{CASE} \ \ \text{OBL}]_1 \\ \text{OBJ}_2 & \\ \text{OBL}_3 & \end{bmatrix}$$

There is a difference in information structure between (9a) and (9b), as, in such constructions, the OBJ realization is discourse-new, unlike what we find with the SUBJ realization (Huddleston and Pullum 2002: 1396–1397). Given this difference in information structure, the two c-structure/f-structure pairs in (17) are not in competition with that in (16), but are in competition with each other. We assume that, in order for two structures to be candidates for optimization, they must be identical in meaning and that a difference in information structure entails a difference in meaning. In both structures in (17), the internal argument of *appear* corresponds to the OBJ; they differ in that there is no SUBJ in (17a) and there is a non-thematic SUBJ in (17b). The optimization for these two candidates is represented in the tableau in (18):

(18) Optimization for (17a) and (17b):

	SUBJCON	GF-FAITH
(17a)	*!	
☞ (17b)		*

As the optimization above illustrates, (17b) satisfies SUBJCON and, although it violates GF-FAITH, it is chosen over (17a), resulting in the grammatical sentence with an expletive subject, i.e., ***There*** *appeared a bird on the windowsill*.

Within this theory, expletive *there* can only be used when it is required to satisfy SUBJCON, as it violates GF-FAITH because it is not lexically required. This includes sentences in which the expletive is in Spec of IP, as in (9b) or (10b), as well as raising to object constructions, where the expletive is in the object position, rather than in Spec of IP, but satisfies SUBJCON in the complement clause, as in (19a):

(19) a. Kim believed there to be flowers in the yard.

b. * Kim believed Ø to be flowers in the yard.

(19a) is chosen over (19b) for the following reason. We assume that raising-to-object verbs such as *believe* or *expect* specify in their lexical entry (but not in their a-structure) that they take a direct GF that is not an argument of the verb. This non-thematic GF is the one that is structure-shared with the subject of the clausal complement of the raising-to-object verb. The expletive subject in the embedded f-structure of (19a) satisfies SUBJCON but violates GF-FAITH, because it is not lexically required by *be*. With respect to the raising object in the main clause, it does not violate the GF-FAITH, since it is lexically required by *believe*. By contrast, the f-structure of (19b) violates SUBJCON in the embedded clause and GF-FAITH in the main clause, as it does not have an expletive subject in the embedded clause and a raising object in the main clause, respectively. The two f-structures and the optimization are represented as in (20) and (21), respectively:

(20) a. F-structure of (19a):

$$\left[\begin{array}{ll} \text{PRED} & \text{'believe} < \mathrm{E}_1,\ \mathrm{I}_2 > \text{'} \\ \text{SUBJ} & [\text{PRED}\ \ \text{'Kim'}]_1 \\ \text{OBJ}_3 & \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'be} < \mathrm{I}_4,\ \mathrm{A}_5 > \text{'} \\ \text{SUBJ} & [\text{CASE}\ \ \text{OBL}]_6 \\ \text{OBJ} & [\text{PRED 'flower'}]_4 \\ \text{OBL} & \left[\begin{array}{ll} \text{PRED} & \text{'in} < \mathrm{I} > \text{'} \\ \text{OBJ} & [\text{PRED}\ \ \text{'yard'}] \end{array}\right]_5 \end{array}\right]_2 \end{array}\right]$$

b. F-structure of (19b):

$$\left[\begin{array}{ll} \text{PRED} & \text{'believe} < \mathrm{E}_1,\ \mathrm{I}_2 > \text{'} \\ \text{SUBJ} & [\text{PRED}\ \ \text{'Kim'}]_1 \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'be} < \mathrm{I}_3,\ \mathrm{A}_4 > \text{'} \\ \text{OBJ} & [\text{PRED}\ \ \text{'flower'}]_3 \\ \text{OBL} & \left[\begin{array}{ll} \text{PRED} & \text{'in} < \mathrm{I} > \text{'} \\ \text{OBJ} & [\text{PRED}\ \ \text{'yard'}] \end{array}\right]_4 \end{array}\right]_2 \end{array}\right]$$

(21) Optimization for (20):

	SUBJCON	GF-FAITH
☞ (20a)		*
(20b)	*!	*

In summary, the distribution of expletive *there* follows from the lexical entry of this expletive and independently motivated constraints. Its occurrence need not be stipulated in the argument structure (or in any other part of the lexical information) of the predicate that cooccurs with it.

4 Analysis of propositional expletive *it*

In this section, we give an account of propositional expletive *it*. We argue against the idea of a special argument-to-function mapping principle licensing a propositional argument as a complement of the verb, in 4.1. In 4.2, we discuss the grammatical function of the propositional constituent in structures with propositional expletive *it*. In 4.3, we give our proposal about propositional expletive *it*. 4.4 highlights the main features of this proposal and compares it with that of Kim and Sag (2005).

4.1 Against a special propositional mapping principle

Propositional *it* can only occur in sentences with a propositional constituent. We propose that propositional *it* is not restricted to satisfy SUBJCON, unlike expletive *there*. One might be tempted to assume that there is a mapping constraint which allows propositional elements to map onto a non-subject function, thereby vacating a position that can be filled by this expletive. We argue that there are three reasons against this assumption. First, "proposition" is not an a-structure notion: it is not a thematic role, but a type of semantic content that an argument may have. An argument role can be semantically a "thing" or a "proposition", among other types (see Jackendoff 1990: 49 for more information about semantic types or conceptual categories), as we see with the subject of *surprise* in (22):

(22) a. That you won the lottery surprised me.
b. The result surprised me.

Thus, there should not be two a-structures depending on whether the theme argument of *surprise* is a "thing" or a "proposition". Likewise, the theme argument of *expect* can also be a thing or a proposition, as in (23), and there should not be two a-structures of *expect* for these two cases. There is just one a-structure with an experiencer and a theme:

(23) a. Nobody expected anything of me.
b. Nobody expected (it) of you that you could be so cruel.

(Kim and Sag 2005: 194)

Second, it would be very problematic to assume that there is a mapping principle that allows a propositional argument to map onto a non-subject GF (whether it is OBJ, as in a framework with a reduced inventory of GFs, such as the present one, or COMP, XCOMP, etc., if a larger inventory of GFs is available), overriding the more general mapping principle that maps the higher of two core arguments to SUBJ (which is implemented in different ways depending on the particular theory, from Bresnan and Kanerva 1989 to Kibort 2001 and other work). For example, *surprise* is a verb whose a-structure includes a cause, which can be a thing or a proposition, and an experiencer argument, expressed as an object, as (22) illustrates. The canonical mapping of arguments to GFs for this verb is represented in (24):

$$\begin{array}{lllll} & & \text{(cause)} & \text{(exp)} & \\ & & | & | & \\ (24) & \textit{surprise} < & \text{E} & \text{I} & > \\ & & | & | & \\ & & \text{SUBJ} & \text{OBJ} & \end{array}$$

If we employed the special propositional mapping rule mentioned, which would allow the propositional *cause* argument to map onto an OBJ/COMP, there should be no reason to insert expletive *it* to fill the SUBJ function, (i.e., it would not be required to satisfy SUBJCON) because the *experiencer* argument would be forced to fill the SUBJ, resulting in the ungrammatical **I surprised that you won the lottery*, with the same meaning as (22a).

Third, as is well known since Postal and Pullum (1988), propositional *it* is not restricted to filling the SUBJ function: it may also be an OBJ, for instance. We repeat examples (6b) and (8) as (25) for ease of consultation:

(25) a. **It** is likely that Susan will be late.
  b. I resent **it** greatly that you didn't call me.
  c. I regret **it** very much that we could not hire Mosconi.

We therefore conclude that it is inadequate to propose special mapping constraints to license the realization of a propositional argument as a complement in order to capture the distribution of propositional expletive *it*. We will assume that the mapping of arguments onto GFs is independent of their semantic type as things or propositions, while looking for other ways of explaining the distribution of this expletive.

4.2 Syntactic function of the extraposed clause

Some words need to be said about the grammatical function of the subordinate clause in examples such as (25). This subordinate clause has been analyzed as an ADJ (i.e., as an appositive clause, e.g. Vikner 1995: 241), a COMP (e.g. Kaplan and Zaenen 1995), or has been suggested to contribute to the same argument slot as the expletive *it*, thus unifying their f-structure information under the same but "discontinuous" function (either SUBJ or OBJ), as proposed by Berman et al. (1998) for German.

We will not assume the discontinuous analysis, for the following reason. In a structure like (25a), the propositional *it* and the subordinate clause would be analyzed as contributing together to the SUBJ function of the matrix clause according to the discontinuous approach. Note that grammatical functions in English are strictly constrained by their c-structure realization: the subject always occupies the specifier position of IP, whereas the canonical object position is VP-internal.[6] Now, while it is true that the propositional *it* occupies the subject position in (25a), and thus is uncontroversially the subject of the clause, the subordinate clause is clearly not occupying the specifier position of IP. Therefore, analyzing the clausal complement as (part of) the SUBJ function would not be appropriate. On the other hand, considering this element as a non-subject function will not cause such a problem.

Adopting the reduced inventory of grammatical functions proposed in Alsina (1996a) and Patejuk and Przepiórkowski (2016) consisting only of SUBJ, OBJ, and OBL (see also Alsina et. al. 2005 and Forst 2006 for arguments against the GF COMP), and taking into consideration the arguments from Kim

[6] The default object position is immediately following the verb, but certain conditions of the object phrase (heaviness, focus or discourse-newness, or a non-nominal category) allow or favor having other sister constituents linearized before it.

and Sag (2005: 197) against analyzing the clausal phrase in extraposition as an adjunct, we will not analyze it as ADJ or COMP. Instead, we assume that the syntactic function of the clausal phrase is always OBJ, which is a distinct function from the function of propositional *it*.

One might object to the claim that the clausal complement in sentences like (25b,c) is an OBJ with the argument that clausal adjuncts cannot appear before objects (pointed out to us by Bresnan p.c.), and therefore the clausal constituent in sentences such as (25b,c) should not be analyzed as bearing the OBJ function because it is preceded by a clausal adjunct such as *greatly* or *very much*. We admit that an NP object cannot be separated from its preceding verb by a clausal adjunct (excepting instances of heavy NP shift), as illustrated below:

(26) a. * She saw often Tom.
b. She saw Tom often.
c. She often saw Tom.

(Huddleston and Pullum 2002: 247)

But when a construction has a clausal object, it is in fact perfectly acceptable for an adverb of the matrix clause to precede the object:

(27) I regret deeply volunteering to take part.

(Huddleston and Pullum 2002: 781)

It can be argued that *regret* takes an object, which can be an NP, as in *I regret my intolerance*, or a clause, as in (25c) or (27). The same happens with three-place predicates, such as *tell*, whose second object can be either an NP or a clause. With two NP objects, it is not possible for an adverb such as *yesterday* in the following example, which modifies the matrix clause, to precede either of the NPs:

(28) a. * He told me yesterday the story.
b. * He told yesterday me the story.

By contrast, the adjunct *yesterday* can appear between the two objects, when the object following the adverbial adjunct is a clause, as opposed to an NP:

(29) He told me yesterday you wanted it.

(Huddleston and Pullum 2002: 781)

As we can see from the example above, *yesterday*, as an unambiguous adjunct of the matrix clause, appears before the clausal object of (29), a double object construction (for the object status of the clausal complement of *tell*, see Dalrymple and Lødrup 2000). Therefore, constructions with an adjunct separating two objects are not necessarily bad: they are only bad if the adjunct appears immediately before an NP object, but not if it precedes a clausal object.

Independent evidence for the assumption that the clausal complement is syntactically an OBJ comes from the impossibility of extraposing the second object of a predicate that takes two objects, as (30) illustrates:

(30) * He told me it that he has tried.

The ungrammaticality of (30) can be explained by analyzing the extraposed complement as an OBJ. Unlike other languages, which allow more than two objects, for example, some Bantu languages (see Bresnan and Moshi 1990, Alsina 1996b, among others), it is clear that English allows at most two objects. The principle that enforces this limitation explains the ungrammaticality of (30), as there would be three objects (i.e., *me*, *it*, and *that he has tried*) if we assume that the extraposed clause is an object. If we assumed the extraposed clause to bear a GF other than OBJ, there would only be two objects in (30), which would leave the ungrammaticality of (30) unexplained.

4.3 Propositional *it*

We assume that the distribution of propositional expletive *it* depends on a special lexical entry that allows it to be used in the presence of an OBJ with propositional semantics:

(31) Lexical entry of propositional *it*:

$$\textit{it}\text{:}\quad \text{NP}_1 \quad \begin{bmatrix} \text{GF} & \begin{bmatrix} (\text{PRED `pro'}) \\ \text{PERS}\ \ 3 \\ \text{NUM}\ \ \text{SG} \end{bmatrix}_1 \\ \text{OBJ}_2 & \end{bmatrix} \qquad \begin{array}{c} \text{Semantic Structure} \\ \begin{bmatrix} \text{TYPE} & \text{proposition} \end{bmatrix}_2 \end{array}$$

The lexical entry in (31) does not restrict the c-structure realization of the propositional complement: it can be a *that*-CP, as in the examples given, a *for-to*-infinitive clause, or a *to*-infinitive clause, etc. For example:

(32) a. It is advisable for students to prepare for the exam.

b. It is important to buy a lottery ticket.

This lexical entry allows propositional *it* to be used in two different situations: the non-thematic (or true) expletive *it* (as in (2)) and the argumental expletive *it*, found in extraposition, as in (25) and (32). The former violates GF-FAITH, as it is not lexically required, but satisfies SUBJCON, very much like expletive *there*; the latter satisfies GF-FAITH, as it is an argument of the predicate, like most NPs, and has the effect of licensing a clausal complement. Let us consider the two in turn.

Non-thematic expletive *it*

On the one hand, *it* can be used in constructions like (2), in which the verb has a single argument with propositional content that maps onto a non-subject function. This is termed the impersonal construction in Huddleston and Pullum (2002: 960) and it involves verbs such as *seem*, *appear*, *happen*, *turn out*, etc.[7] With these verbs, the subordinate clause cannot appear in subject position, but only in postverbal position, which distinguishes them from extraposable predicates such as *surprising*, which can have a clausal expression either preverbally or postverbally:

(33) a. It seems that he was lying.

b. * That he was lying seems.

[7] These verbs also have a use as raising verbs, taking a predicative complement, instead of a full clausal complement, but we will not be concerned with this use here.

(34) a. It is surprising that he was lying.

b. That he was lying is surprising.

We therefore propose that predicates of the impersonal construction like *seem* lexically specify both the grammatical function and the grammatical category of their single argument: it is a CP and an OBJ:

(35) *seem* < Arg >
$$\begin{array}{c} \text{< Arg >} \\ | \\ \text{OBJ} \\ | \\ \text{CP} \end{array}$$

Given that the sole argument of *seem* maps onto an object, the clause in which this verb appears needs a subject. In this situation, propositional *it* fills the non-thematic SUBJ function, satisfying SUBJCON. Notice that this GF is not lexically required, and thus violates GF-FAITH; but since SUBJCON ranks higher than GF-FAITH in English, the structure with the expletive subject is chosen over the one that lacks it. Being a non-thematic subject, the option without PRED is chosen. In such cases, the subject *it* is a true, or non-thematic, expletive.

Extraposition *it*

On the other hand, the lexical entry in (31) allows any argument that can be expressed as an NP and is semantically consistent with a proposition to be expressed by means of *it*, which licenses a clausal object that provides the propositional content of the argument. This is the extraposition *it*, which we find in (25) and (32). In these examples, *it* is not a true expletive – if we take *true expletive* to be a GF that does not map onto a semantic participant – as it maps onto an argument of the predicate. In these cases, the option with PRED is chosen. The GF with propositional content does not correspond to an argument of the verb, but yet satisfies GF-FAITH, as it is required by the lexical entry of *it*.

Thus, a single lexical entry for expletive *it*, in (31), in combination with the general constraints GF-FAITH and SUBJCON, gives rise to both the non-thematic expletive *it* of impersonal constructions, as in (33), and the argumental expletive *it* of the extraposition construction, as in (25), (32), and (34). A consequence of this theory is that, in the extraposition construction, expletive *it* can only occur in the position that corresponds to the argument GF it realizes, whether SUBJ or OBJ, whereas the phrase that corresponds to the GF with propositional content appears in postverbal position, as a clausal object, which explains the ungrammaticality of (36), as expletive *it* is not in the position that corresponds to its GF, nor is the phrase with propositional content.

(36) * That you won the lottery surprised me **it**.

(cf. (5c))

However, nothing that we have said so far explains the ungrammaticality of (37a), where expletive *it* is an OBJ, as corresponds to the argument it fills, and the phrase with propositional content is also an OBJ, as required by the lexical entry (31). The grammatical example (37b) has the same elements as (37a), only in a different linear order.

(37) a. * I resent that you didn't call me it.
b. I resent it that you didn't call me.

Following Kim and Sag (2005), we can assume the existence of a linear precedence rule that requires clausal phrases to linearly follow a sister GF:

(38) Linear precedence rule 1:
A clausal phrase must follow a sister GF.

As the two postverbal constituents in (37) are OBJ and they are sister constituents, the clausal object must follow the NP OBJ *it*, which explains the contrast between (37a) and (37b).

Another fact that needs to be explained is that, as observed by Huddleston and Pullum (2002), clausal extraposition is normally required when there is an object predicative complement (such as *hardly surprising* in (39)), except if the clause is topicalized (as in (39c)):

(39) a. * I find that he tried to retract his statement hardly surprising.
b. I find **it** hardly surprising that he tried to retract his statement.
c. That he tried to retract his statement I find hardly surprising.
(Huddleston and Pullum 2002: 963)

The ungrammaticality of (39a) follows from the linear precedence rule (38). However, reordering of the predicative phrase and the extraposed clause yields another ungrammatical, or marginal, sentence:

(40) * I find hardly surprising that he tried to retract his statement.

We can explain this fact by adopting the explanation in Kim and Sag (2005), which involves another linear precedence statement that requires a subject to precede the phrase of which it is the subject

(41) Linear precedence rule 2:
If XP is the subject of YP, XP linearly precedes YP.

Example (40) does not comply with this precedence rule, as the subject of *hardly surprising* – the *that*-clause – follows that phrase, instead of preceding it. In contrast, example (39b), with expletive *it*, meets both precedence requirements, as the *that*-clause follows its sister predicative complement, as well as *it*, in compliance with linear precedence rule 1, and *it* precedes the predicative phrase of which it is the subject, in compliance with linear precedence rule 2.

A last fact that needs to be explained regarding the extraposition construction is that the propositional constituent in this construction cannot be topicalized:

(42) * That you won the lottery, **it** surprised me.

The ungrammaticality of topicalizing an extraposed complement clause does not mean that a complement clause cannot be topicalized at all. It can be preposed, without propositional *it*, as in (39c), or in the following example (43):

(43) For them to sack him we would regard as a gross miscarriage of justice.
(*cf*. We would regard it as a gross miscarriage of justice for them to sack him.)

(Huddleston and Pullum 2002: 1255)

The same observation is also reported by Kaplan and Zaenen (1995) and Kim and Sag (2005), among others:

(44) a. That Susan would be late John didn't think was very likely.
b. * That Susan would be late John didn't think **it** was very likely.

(Kaplan and Zaenen 1995: 158)

(45) a. That Kim would lose to Pat, nobody had expected.
b. * That Sandy snores, **it** bothers Kim more and more.

(Kim and Sag 2005: 202)

Note that informants providing judgement about (42) point out that the sentence is acceptable. They nevertheless also point out that the sentence would be more natural if *it* is replaced by *that*, i.e.:

(46) That you won the lottery, **that** surprised me.

The possibility of using a demonstrative *that* shows that the acceptable utterance *That you won the lottery, it surprised me* is a case of left-dislocation, instead of topicalization of an extraposed clause; and the *it* is not an expletive *it*, but a pronominal *it* with explicit reference. In the unacceptable situation of (42), *it* is the expletive, which indicates that it is ungrammatical to topicalize an extraposed clause. The explanation that Kaplan and Zaenen (1995) propose for this resorts to a restriction on functional control in long-distance dependencies. They assume that the extraposed clause bears the function COMP and that the functional uncertainty equation that models long-distance dependencies cannot have the GF COMP as its bottom. In other words, they assume that a COMP cannot undergo topicalization.[8] But see Alsina et al. (2005) for an alternative analysis that does not involve COMP.

Our explanation of this last fact, i.e., the ungrammaticality of topicalizing an extraposed clause in English, takes into account the relation between clausal heaviness and extraposition. We are in line with Huddleston and Pullum (2002: 1403) in considering that, "the effect of extraposition is to place a heavy constituent at the end of the clause". Let us assume that expletive *it* marks the clausal object in its lexical entry as heavy, [H+], and that [H+] constituents are linearized as final in their clause. It follows from this that the clausal constituent in a clause with expletive *it* cannot be preposed (i.e., topicalized), as then it would not be final. In contrast with the clausal object of *it*-clauses, other clausal objects are not constrained to be [H+] and therefore are free to be preposed (i.e., topicalized), as in (39c) or (43).

[8] See Berman et al. (1998), who report a similar contrast in German between the ungrammaticality of topicalizing an extraposed clause and the possibility of topicalizing a clausal complement and propose an explanation that depends on the assumption that the expletive and the extraposed clause are coarguments.

To conclude this subsection, we would like to point out a consequence of our analysis of propositional expletive *it*. Even though the same lexical entry licenses both extraposition *it* and the non-thematic dummy *it*, these two uses of the expletive have a different representation, as has been noted: extraposition *it*, being an argument, has the [PRED 'pro'] feature, whereas the non-thematic expletive lacks this feature. This difference implies that expletive *it* cannot be the shared constituent in a coordination of an impersonal and an extraposition construction, as shown in (47):

(47) a. It seemed that he was trying to hide his true identity.
b. It was later confirmed that he was trying to hide his true identity.
c. * It seemed and was later confirmed that he was trying to hide his true identity.

(Huddleston and Pullum 2002: 961)

According to Huddleston and Pullum (2002: 961), the ungrammaticality of (47c) suggests that the extraposed clause "does not have the same function in the two cases" (i.e., in (47a) and (47b)). However, according to our analysis, the *that*-clause does have the same function in both cases, namely, object, but the ungrammaticality of (47c) is attributed to the impossibility of expletive *it* being at the same time non-thematic in the *seem* case, as in (47a), where it lacks a PRED value, and thematic in the extraposition case, as in (47b), where it has a PRED value. The following example shows that expletive *it* cannot be the shared constituent in a coordination of an impersonal construction and an extraposition construction also when each conjunct contains its own clausal complement:

(48) * It seemed that he didn't stand a chance and was hardly surprising that he didn't win.

The ungrammaticality of this example cannot be attributed to the putative difference in the grammatical function of the clausal complement in the two constructions involved, but can be attributed to the incompatible requirements imposed in each construction to the expletive *it*.

4.4 Comparison with Kim and Sag (2005)

Kim and Sag (2005), henceforth KS, develop an analysis of the English extraposition construction within HPSG that has some similarities with the present proposal. They propose a rule that creates new words out of words whose SUBCAT feature includes an S or CP argument such that in the new words this S or CP argument is not in the SUBCAT feature, but in the EXTRA feature. Also, an expletive NP (NP[*it*]) holds the place of the extraposed complement in the new word's SUBCAT list.

We will not analyze here the advantages or disadvantages of introducing the additional selection feature EXTRA, although there seems to be little independent evidence for it. The differences between the KS analysis and the present one that we will focus on are: (a) the difference in generality between the two analyses; (b) the difference regarding verbs that select a complement that is necessarily clausal; and (c) the difference regarding verbs that select a complement that is necessarily an NP.

Whereas in the present theory the same lexical entry for expletive *it* accounts for the use of this expletive in both the impersonal construction and the extraposition construction, the KS analysis limits its scope to the extraposition construction. It is clear that the two constructions have common elements: in both cases, the same expletive is used and the structure includes a clausal complement. This is captured in the present theory, but no indication is given that the KS analysis of the extraposition construction can be extended to the impersonal construction. We will not speculate as to whether and how this can be done, but, while the lexical rule approach implies that the rule applies optionally, the impersonal construction has the subject expletive *it* and the complement clause as obligatory elements.

The two analyses make differing predictions with respect to the classes of verbs that can appear in the extraposition construction. For KS, the extraposition rule is only possible with verbs that select an argument that can be expressed as a clause and, therefore is not possible with verbs whose arguments are restricted to be of other categories, such as NP or PP. In contrast, the present analysis restricts extraposition to occur with verbs that select an argument that can be expressed as an NP: the expletive *it*, being an NP, can only appear in positions where NPs are possible; the argument must also allow the propositional semantics associated with the clausal object licensed by *it*.

With respect to verbs that take an argument that is constrained to be an NP, the present theory predicts that this argument should be expressible by means of extraposition *it*, whereas the extraposition rule of KS cannot be used with it. These verbs include *take*, *put, like*, or *dislike*, etc. With such verbs, a CP complement without the co-appearance of extraposition *it* will result in an ungrammatical construction. For example, the object of *put* is obligatorily an NP, and appearance of a CP object is acceptable with this verb, provided that expletive *it* also appears. On the contrary, without extraposition *it*, such sentences are judged to be unacceptable. The contrast is illustrated as follows:[9]

(49) I put **it** to you that you know what the consequences would be.
(*cf.* * I put to you that you know what the consequences would be.
* I put that you know what the consequences would be to you.)
(Huddleston and Pullum 2002: 247)

As for verbs that only take CP – but not NP – complements, the present analysis predicts they do not allow extraposition *it*, whereas KS lead us to expect them in the extraposition construction. These verbs include *object*, *conclude*, *reason*, *reflect*, *reply*, *complain*, *decide*, etc. Some of these verbs cannot take any kind of complement except for a clause, as is the case of *reason*

[9] KS account for these cases by means of additional lexical entries of the verbs involved that directly stipulate that they take the expletive *it* as an object and a clausal structure in the EXTRA feature.

(but not *reason out*, which does take an NP object). Others have alternative lexical entries taking either an NP, an oblique PP, or a CP complement. For example, *conclude* can take either an NP object or a CP complement, with different meanings, as in *I concluded my work* (where *conclude* equals *finish*) and *I concluded that there was no satisfactory solution* (where *conclude* equals *reach the conclusion*), respectively. *Complain* can take either an oblique *about*-PP, which might also be analyzed as an adjunct, or a clause. In *I complained about the weather*, *the weather* is the object or target of the complaint; in *I complained that it was too hot*, the CP is the argument on which the complaint is based. With all of these verbs when taking an CP complement, extraposition *it* is bad:

(50) a. Local authorities complained (*it) that they lacked sufficient resources.

(based on Oxford Dictionary of English, p.356)

b. The boy's father objected (*it) that the police had arrested him unlawfully.

(based on Oxford Dictionary of English, p.1226)

These facts confirm our prediction that verbs that take a complement that can only be expressed as a CP will not allow extraposition of this complement and constitute a crucial bit of evidence to distinguish our analysis from KS. In their discussion of different classes of verbs with respect to extraposition, KS only mention *think* as an example of Group III verbs, i.e., verbs that only take a clausal complement, but not an NP object. However, this verb turns out not to be a very good example of this type, because it also takes NP objects in restricted conditions, as in *What are you thinking?* or *What do you think?*, given that *what* is an NP, or *She thought something else entirely* or *She thought a few things I cannot explain*, etc. This shows that *think* can take an NP object. Consequently, just as the NP object of *think* is possible, with semantic limitations, it is not surprising that extraposition *it* is possible with this verb, although not at all frequent:

(51) I thought it that it would be nearly impossible for the filmmakers to sustain such a level of excitement through the rest of the movie.

(Kim and Sag 2005: 209)

To conclude this subsection, the present theory not only accounts for the distribution of expletive *it* in both the impersonal and the extraposition construction by means of a single lexical entry, but correctly predicts that, in the extraposition construction, this expletive can appear wherever we have a verb that takes an NP subject or object and cannot appear in place of the complement with verbs whose complement is constrained to be a CP. KS make the opposite prediction: they predict extraposition to be possible with the latter class of verbs and impossible when the expletive corresponds to an argument whose categorial expression is constrained to be NP.

5 Conclusions

In this paper, we have proposed a theory in which expletives are not stipulated in the lexical entry of the predicate that cooccurs with the expletive, but their distribution instead follows from general principles such as the Subject Condition and from the lexical entries of the relevant expletives. As a result, there are no expletive insertion rules or lexical rules to generate verbs that select expletives, no alternative lexical entries for verbs depending on whether they have an expletive or not, and no need to stipulate in any way the FORM feature of the expletives.

An idea that the present theory crucially depends on is that the assignment of GFs to arguments and the licensing of GFs in a clause take place in the syntax. Whereas standardly in LFG this process is assumed to take place in the lexicon, so that words exit the lexicon with the list of GFs that they take, we assume that the lexical entries of predicates do not fully specify the GFs that they take, but in general only specify the argument structure of the predicate, which constrains, but does not determine, the GFs associated with the predicate. In other words, argument structure in our paper replaces the lexical form in previous analyses, and there is no list of GFs in the PRED value. This is necessary for two reasons. First, the Subject Condition plays an important role in licensing the expletives *there* and non-thematic *it*: the Subject Condition is a constraint on f-structures and it helps choose f-structures with a subject over f-structures without a subject, even if that subject is not an argument of the predicate. Second, the complement clause in extraposition is licensed by expletive *it*; the predicate that cooccurs with that clause should not foresee in its lexical entry that it takes a complement clause; rather, if one of its arguments can be an NP and is realized as the expletive NP *it*, it is this word that licenses the complement clause, thanks to GF-FAITH, a reinterpretation of Coherence.[10]

The proposed analysis indicates that there are two kinds of expletives: true, or non-thematic expletives, as the case of *there* and non-thematic *it*, and argumental expletives, as the case of extraposition *it*. Our proposal about the two expletives, especially expletive *it*, makes use of a reduced inventory of grammatical functions: the three strongly motivated SUBJ, OBJ and OBL, as argued by Alsina (1996a), Alsina et al. (2005), Forst (2006), and Patejuk and Przepiórkowski (2016); and we do not need to enrich the inventory with other grammatical functions, such as COMP or XCOMP, as many LFG analyses such as Kaplan and Zaenen (1995) or Bresnan (1982), etc. do, or other theoretical constructs, such as EXTRA, as in Kim and Sag (2005), which do not have strong independent motivation.

[10] The effects of Completeness are captured by the assumption that the argument-to-GF mapping principles apply whenever possible and that lexical requirements must be satisfied in the syntax.

References

Alsina, Alex. 1996a. *The Role of Argument Structure in Grammar: Evidence from Romance*. Stanford: CSLI Publications.

Alsina, Alex. 1996b. Passive types and the theory of object asymmetries. *Natural Language & Linguistic Theory* 14(4), 673–723.

Alsina, Alex, K. P. Mohanan, and Tara Mohanan. 2005. How to get rid of the COMP. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, 21–24. Stanford: CSLI Publications.

Berman, Judith, Stefanie Dipper, Christian Fortmann, and Jonas Kuhn. 1998. Argument clauses and correlative *es* in German – deriving discourse properties in a unification analysis. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*. Stanford: CSLI Publications.

Bresnan, Joan. 1982. The passive in lexical theory. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 3–86. Cambridge MA: The MIT Press.

Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 70(1), 72–131.

Bresnan, Joan and Jonni Kanerva. 1989. Locative inversion in Chicheŵa: a case study of factorization in grammar. *Linguistic Inquiry* 20(1), 1–50.

Bresnan, Joan and Lioba Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21(2), 147–186.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge MA: The MIT Press.

Dalrymple, Mary and Helge Lødrup. 2000. The grammatical function of complement clauses. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG00 Conference*, 104–121. Stanford: CSLI Publications.

Falk, Yehuda. 2001. *Lexical-Functional Grammar: an Introduction to Parallel Constraint-Based Syntax*. Stanford: CSLI Publications.

Findlay, Jamie. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4(2), 293–338.

Forst, Martin. 2006. COMP in (parallel) grammar writing. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, 222–239. Stanford: CSLI Publications.

Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge MA: The MIT Press.

Kaplan, Ronald and Joan Bresnan. 1982. Lexical-functional grammar: a formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge MA: The MIT Press.

Kaplan, Ronald and Annie Zaenen. 1995. Long-distance dependencies, constituent structure, and functional uncertainty. In Mary Dalrymple, Ronald Kaplan, John Maxwell, and Annie Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, 137–166. Stanford: CSLI Publications. First printed in Mark R. Baltin and Anthony S. Kroch, (eds.). 1989. *Alternative Conceptions of Phrase Structure*, 17–42. Chicago: University of Chicago Press.

Kibort, Anna. 2001. The Polish passive and impersonal in Lexical Mapping Theory. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, 163–183. Stanford: CSLI Publications.

Kim, Jong-Bok and Ivan Sag. 2005. English object extraposition: a constraint-based approach. In Stefan Müller (ed.), *Proceedings of the HPSG05 Conference*, 192–212. Stanford: CSLI Publications.

Krejci, Bonnie. 2014. What is raining? English weather *it* revisited. *Linguistic Society of America* 88, Minneapolis, MN. 3 January.

Kuhn, Jonas. 2003. *Optimality-Theoretic Syntax: a Declarative Approach*. Stanford: CSLI Publications.

Levin, Beth. 2017. Talking about the weather: a case study of precipitation verbs. Linguistics Colloquium. University at Buffalo.

Oxford Dictionary of English. 2010. Ed. by Angus Stevenson. Third Edition. Oxford: Oxford University Press.

Patejuk, Agnieszka and Adam Przepiórkowski. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King and Stefan Müller (eds.), *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 541–559. Stanford: CSLI Publications.

Postal, Paul and Geoffrey Pullum. 1988. Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19(4), 635–670.

Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford University Press.

A One-level Analysis of Icelandic Quirky Case

Avery D. Andrews
The Australian National University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 27–47

http://csli-publications.stanford.edu/LFG/2019

Keywords: Icelandic, DBA glue, description by analysis, split lexicon, quirky case, Kibort-Findlay Mapping Theory

Andrews, Avery D. 2019. A One-level Analysis of Icelandic Quirky Case. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 27–47. Stanford, CA: CSLI Publications.

Abstract

This paper presents a single level analysis of the f-structure of Quirky Case NPs in Icelandic that covers the data of the two-level analysis presented by Andrews (1982, 1990), using the 'split lexicon' and DBA Glue proposal of Andrews (2007, 2008) to deal with the phenomena that motivated the two-level analysis. The resulting analysis is simpler in some ways (although perhaps a bit more stipulative in others), and more consistent with recent developments in LFG such as the Kibort-Findlay Mapping Theory.

1 Introduction

Andrews (1982, 1990) proposed a 'two level' analysis of arguably irregular, or 'Quirky'[1] case in Icelandic in LFG that accommodated a considerable range of difficult data. But there have been both empirical and theoretical developments that indicate that it might be time for a substantial revision.

Empirically, perhaps the most important factor is something that didn't happen with the original analysis: it did not become integrated into a general typology of case-marking and agreement. Its main point was to explain why Quirky Case NPs fail to trigger agreement on certain items which are agreement targets for regularly case-marked NPs, but nevertheless do trigger agreement on certain other kinds of targets, such as secondary predicates. But there are languages like Warlpiri, where inherent case marking has little or no effect on agreement, and Hindi, where regular case-marking suppresses agreement in the same way that arguably non-regular case-marking does. Another empirical point is that the two-layer analysis proposed for agreement with nominative objects doesn't do a great job with the further data presented by Sigurðsson & Holmberg (2008) and later authors such as Ussery (2017a), and, furthermore, doesn't generalize to the data of long distance agreement with object in Hindi from Bhatt (2005).

Theoretically, on the other hand, the two-level hypothesis is not easily compatible with the new Kibort-Findlay Mapping Theory (Findlay, 2016), and neither with any of the previous lexical mapping theories that have been explored in LFG. And there are various technical problems and unexplained phenomena in the data.

A final factor that facilitates a different analysis is the Split Lexicon and DBA Glue proposal of Andrews (2007, 2008), which proposes that the traditional LFG lexicon should be replaced by two, a Morphological Lexicon that

[†]I would like to thank the two reviewers for very extensive and helpful comments.

[1]Nobody knows who invented this term, and there is furthermore a tendency to reserve 'quirky' for accusative and genitive subjects and genitive objects, which are less semantically predictable than the datives (van Valin, 1991), which are then called 'inherent' (Thráinsson, 2007, 181-182).

is very similar to the traditional LFG lexicon, but without any direct involvement with meaning, and a Semantic Lexicon that associates combinations of features (both PRED-features and grammatical features) with meanings, using Glue semantics, but with the meaning-constructors applying (primarily) to f-structures.[2] This allows a workable analysis to be formulated with less stipulation than its most fully worked out unification-based competitor, Sag et al. (1992), which requires two kinds of case attributes, plus provisions to equate them under certain circumstances. With the split lexicon, these provisions can be replaced by the distinction between 'uninterpreted' (structural) and '(co-)interpreted' (lexical/inherent/quirky) case, which has a natural representation in the structures.

In the following sections, I briefly sketch the main features of the original analysis, and discuss some of the theoretical and empirical problems that motivate a revision. Then I describe how the Split Lexicon works, apply it to develop the proposed new analysis, and, finally integrate that with KFMT.

2 Highlights of the Original Analysis

Icelandic, like German, has preserved the four cases of nominative, accusative, genitive and dative, although the endings are more numerous and distinct than in German, and the three genders (masculine, feminine and neuter) are distinguished in the plural as well as the singular. Furthermore, predicate adjectives and passive participles agree with their subjects if these are case-marked in accordance with the regular structural case-marking rules, whereby subjects are nominative, and objects are accusative if the subjects are regularly case-marked[3] (Thráinsson, 2007, 157–159):

(1) a. Hún er rík.
she.NOM is rich.NOM
'She is rich.'

b. Hún var handtekin.
she.NOM was arrested.NOM
'She was arrested.'

c. Þeir segja hana (vera) ríka.
they say her.ACC (to be) rich.ACC.
'They say that she is rich.'

[2]This is 'Description by Analysis' (DBA) rather than the 'co-description' (Halvorsen & Kaplan, 1995) that is normally used for Glue.

[3]The working technical formulation is a bit tricky, and in many situations the dative can be regarded as 'regular' on a semantic basis, but we ignore these issues here. See van Valin (1991), Barðdal (2011a, 2011b) and much further literature for discussion.

d. Þeir segja hana (hafa verið) handtekna.
they say her.ACC (to have been) arrested.ACC
'They say that she has been arrested.'

LFG accommodates these and more complex examples by using the mechanism of functional control to allow one NP to be simultaneously the subject or object of multiple verbs, adjectives etc. at the same time, so that they all agree with it, and case marking is determined by the overt position of the NP.

But Icelandic also has interesting examples of non-agreement with subjects. There are a considerable number of verbs that take subjects in the dative case, fewer that do in the accusative, and a very few in the genitive. There are an impressive number of arguments that these 'putative non-nominative subjects' really are subjects, that is, they function as subjects for a number of grammatical properties, and thereby reside substantially in subject position. Among the strongest of these arguments comes from the fact that they occur covertly, as subjects of infinitives in complements which have empty/PRO/null subjects (Thráinsson, 2007, 159, 165):

(2) a. Stelpunum leiddist í skólanum.
the girls.DAT was bored in school
'The girls were bored in school.'

b. Stelpurnar vonast til að leiðast ekki í skólanum.
the girls.NOM hope towards to be bored not in school
'The girls hope not to be bored at school.'

Unlike in English, Icelandic infinitives in this kind of complement cannot have overt subjects (if the subject is not coreferential with a suitable controller, a subjunctive clause is used), and, furthermore, Icelandic is not really a pro-drop language, so (b) has to be an infinitive with an obligatorily suppressed something, most plausibly identified as a subject, since clear cases of objects cannot be suppressed in this way.

So we can now state the interesting fact, which is that except perhaps in some recent, innovative varieties of the language, finite verbs never agree (in person and number) with their non-nominative subjects, while adjective and passive participles agree (in gender, number, and case) only under certain limited circumstances. Nonagreement in number with a dative subject has already been seen in (2a); nonagreement in person and with accusative subjects are illustrated below (Thráinsson, 2007, 159):

(3) a. Mér býður við setningafræði.
me.DAT loathes.3SG against syntax
'Syntax makes me sick.'

b. Strákana rak á land á eyðeyju.
boys-the.ACC.PL drifted.SG to shore on desert island
'The boys drifted ashore on a desert island.'

Quirky Case is also preserved and fails to trigger agreement under Passive and 'Subject-Raising', as discussed below and extensively in the literature.

Andrews' proposal was that in the f-structure of the Quirky subjects and objects, there is an extra structural layer that both hides the agreement features of the NP from most things that might want to agree with it, and also prevents the regular case-marking rules from applying, and thereby ruling the sentences out by producing contradictions. In order to emulate the no longer very popular $(\uparrow(\downarrow \text{PCASE})) = \downarrow$ analysis from (Kaplan & Bresnan, 1982, 197–202), Andrews used the case-name as a grammatical functional label, although a constant GF such as OBL could also be made to work:

(4) $$\left[\text{SUBJ} \quad \left[\text{ACC} \quad \begin{bmatrix} \text{GEND} & \text{MASC} \\ \text{NUM} & \text{PL} \\ \text{CASE} & \text{ACC} \\ \text{PRED} & \text{`Boy'} \end{bmatrix}\right]\right]$$

So if a verb comes along wanting to require that the NUM-value of its SUBJ be PL, it wouldn't match up with the 'real' number value provided by the noun, but be stuck on the top level, where it will fail, due to agreement values being associated with constraining equations..

Non-Quirky NPs would on the other hand have only a single layer in their f-structure, and the regular case-marking rule was that a first or second object (OBJ, OBJ_θ in the analysis to come) would be marked accusative as long as the subject was also non-Quirky. Nominative was treated as the unmarked case, which allows nominative to be the default case value on an object when the subject is Quirky, which gives reasonable results, including in constructions where a nominative subject is raised into nominative object position, where the embedded object remains accusative, just as it would in a normal accusative plus infinitive (ACI) construction with a nominative matrix subject:

(5) Mér virðist hún hafa þann galla einan, ...
me.DAT seems it.NOM.F to have that flaw.ACC only.ACC
'It seems to me to have only that flaw . . ..'
http://timarit.is/view_page_init.jsp?pageId=4411344

This covers non-agreement, but there are also instances of agreement with Quirky NPs that need to be accounted for. These fall into three types:

(6) a. Secondary predicates

b. Control complements of certain verbs that take Quirky (dative) objects

c. Some complicated examples which I claim to be plausibly performance errors

The secondary predicate exceptions are especially interesting in a strengthened form of the complement subject deletion arguments, indicating that the infinitives really did have nonovert subjects in their usual case, even if that case was Quirky (Thráinsson, 2007, 417):

(7) a. að vanta einan í tíma er vandræðlegt.
to be missing alone.ACC.M.SG in class is embarassing
'It is embarassing to be alone missing from class.'

b. að vera kastað einum í dýflissu er hræðilegt.
to be thrown alone.DAT.M.SG in dungeon is terrible
'It is terrible to be thrown into the dungeon alone.'

Of course, the full sentence versions of the infinitive clauses here with their overt subjects are also fine (Thráinsson, 2007, 416).

Some control complement examples from Andrews (1990) are:[4]

(8) a. Þeir lýstu glæpamanninum sem
they described the criminal.DAT.M.SG as
stórhættulegum.
very dangerous.DAT.M.SG
'They described the criminals as very dangerous.'

b.
Glæpamönnum var lýst sem stórhættulegum.
the criminals.DAT.PL was described.SUP as very dangerous.DAT.PL
'The criminals were described as very dangerous.'

c. Hann heldur tönnunum sínum hvítum og hreinum.
he keeps teeth his.DAT.PL white.DAT.PL and clean.DAT.PL

Note that in (b), the adjective agrees with the dative subject while the passive auxiliary and participle do not.

The explanation for the agreement with the secondary predicates and dative-controlled complements that was presented in Andrews (1990) was that in order for the results of secondary predication to be semantically interpretable, the secondary predication rule would have to set the inner structure of the NP

[4]SUP represents 'supine', a form that is morphologically nominative/accusative neuter singular

rather than the entire structure as the SUBJ-value of the adjective, rendering the agreement features visible to agreement. And likewise for the control complement examples, except that it would be the control equations associated with the matrix verbs that did this. This is workable, although it does lead to the implication that there could be languages where Quirky NPs could not be subjects of secondary predicates, which to the best of my knowledge has not been documented.

3 Problems with the Two-Level Analysis

All of this worked reasonably well, in spite of some technical issues, but various problems either emerged over the decades, or were not cleared up. We discuss some but not all of them here, while another, integration with KFMT, will be discussed later when we explain how that integrates with the one-layer analysis.

Perhaps the most serious is that the analysis does not seem to have found a clear place in any reasonable typology of the interactions of case and agreement. The simplest expectation from the analysis would be that the lexically controlled case inhibits agreement, while regularly controlled case does not, but this is false. Warlpiri for example has lexically controlled ergative on subjects, and dative on objects, but the former has no effect at all on agreement (person-number marking on auxiliaries), and the latter hardly any, and that is furthermore enhancing: an overt clitic rather than null for a dative object, as originally noted by Hale (1973), with later supporting argumentation about the grammatical relations and related phenomena by Simpson & Bresnan (1983) and Simpson (1991). Using a two-level analysis for lexically controlled agreement in Warlpiri[5] would require complexifying the conditions for both subject and object agreement.

The opposite problem is provided by Hindi, where Butt & King (2003) discuss in some detail, in an LFG framework, how non-lexically controlled case-marking with a combination of semantic and structural conditioning on both subjects (ergative *ne*) and animate or definite objects (*ko*) inhibits agreement completely. Technically, this can be easily handled by limiting these agreements to nominative case triggers, but the more general point is the absence of a typology where case-marking implemented by two levels plays a clear role.

Another relevant issue is a decline in potential theoretical support from other directions for the two-level analysis. Before the advent of LFG's Glue semantics in the early mid 1990s, it seemed plausible to claim that semantically case-marked NPs needed an extra structural level for a PRED-feature. So a sentence such as (9) might get a structure like (10):

[5] Exemplified by some intransitive verbs that take ergative subjects, and transitive verbs that take dative objects; a survey of case-marking patterns of verbs is given in Hale (1982).

(9) ngatju pirli-ngka.
I.ABS hill-LOC
I am on the rock/hill (Warpiri, Simpson, 1991, 215)

(10) $\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Pro'} \\ \text{PERS} & \text{I} \end{bmatrix} \\ \text{PRED} & \text{`Loc(SUBJ, OBJ)'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`Rock/Hill'} \end{bmatrix} \end{bmatrix}$

In particular, the rock/hill is introduced into the f-structure as the OBJ of its locative case-marker, which, among other things, averts the possibility of (9) being interpreted as 'I am a/the rock/hill'. For more discussion see Simpson (1991, 196, 215).

But Glue semantics[6] changes this, by allowing grammatical features to appear in a flat structure, but nevertheless introduce operators that apply semantically in succession. This is can happen because an inflected form can introduce a meaning-constructor that in effect operates on the meaning currently associated with an f-substructure and provides a new one. A possible analysis for a locative case might therefore be:

(11) $\lambda yx.At(x,y) : \uparrow_e \to (\uparrow_e \to \uparrow_t)$

This converts an entity (corresponding to the first argument) into a predicate over entities (corresponding to the following two arguments) that is true if and only if the second argument entity is located *At* the first argument entity.[7] Space does not permit elevating this to a full analysis, but something using PREDLINK (Laczkó (2012) and references cited there) seems plausible, to keep the locative NP's f-structure distinct from that of the sentence:

(12) $\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Pro'} \\ \text{PERS} & \text{I} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`Rock'} \\ \text{CASE} & \text{LOC} \end{bmatrix} \end{bmatrix}$

Because Glue assembly can do the work of the Completeness and Coherence constraints, it is not even necessary for the entire structure to have a PRED of its own, although analyses using PREDLINK tend to assume this.

[6]See Dalrymple (2001), Asudeh (2005), Andrews (2010) and Asudeh (2012) for presentations of Glue, Andrews using a somewhat different presentation than Dalrymple and Asudeh, although other than the absence of a semantic projection in Andrews, the theory is the same.

[7]A semantic projection is not used here, because the semantic projection is not needed for this analysis.

That two-level analyses were motivated for semantically case-marked NPs does not imply that they were available for argument NPs, but it does make such an analysis more plausible, since the machinery for generating such NPs would have to be available in the absence of obvious overt evidence for the structures, as can be found for locative NPs in Bantu languages (Bresnan & Mchombo, 1995).

There have been further developments in LFG since the 1980s which make it easier to capture all of the original data without a two-level analysis. One of these is the concept of 'inside-out-functional uncertainty' (iofu), which makes it possible to write into a lexical item a constraint meaning 'I am an adjunct', such as:[8]

(13) (ADJUNCT ↑)

The idea here is that '↑' designates the f-structure that the item is appearing in, and the sequence of grammatical functions in front of ↑ indicates a list of grammatical functions which one must be able to climb up, in inverse order, from that f-structure. See Nordlinger (1998) for discussion with a focus on case-marking in Australian linguistics. We will return to this when we need to use it.

Another problem arises with the phenomenon of agreement with nominative objects. Recent work on this has been reviewed and extended by Ussery (2017a), but Andrews (1990, 211–213) discusses a form of example that does not seem to have been much considered in the literature, with at least some exceptions, such as Alsina & Vigo (2017).[9] These are cases where a matrix verb agrees optionally with the object of its functionally controlled (ECM) complement:

(14) a. Honum eru taldir hafa verið gefnir
him.D are believed.M.N.PL to have been given.M.N.PL
hestarnir.
the horses.M.N.PL
'To him are believed to have been given the horses.'

b. Honum er talið hafa verið gefnir
him.D is believed.SUP to have been given.M.N.PL
hestarnir.
the horses.M.N.PL
'To him are believed to have been given the horses.'

[8]Such expressions are in general instances of functional uncertainty, because f-structures can contain re-entrancies.

[9]Who cites Sigurðsson (2004), where I can't this form of example, although I do think I recall that he has discussed them somewhere.

c.	*Honum er talið	hafa	verið	gefnið	hestarnir.	
	him.D is believed.SUP	to have	been	given.SUP	the horses.M.N.PL	

In (a) we see agreement of the matrix verb with the complement object, in (b) nonagreement, while in (c) we see the ungrammatical result of neither complement nor matrix verbs agreeing. This is a consequence of the fact that agreement of a passive with its nominative second object is obligatory, even though most other cases of agreement with nominative objects are optional. The judgements from a questionnaire returned by seven people at Háskoli Íslands are:[10]

(15)

	✓	?	?*	*?	*
a)	4	2	0	0	1
b)	5	1	1	0	0
c)	0	1	0	0	6

The questionnaire results justify treating (a) and (b) as grammatical, and (c) as ungrammatical, although Alsina and Vigo, working in an OT framework, claim only (a) to be grammatical.

Andrews' proposal was that there was an equation in lexical entries that would copy the gender, number and case of a nominative object to the outer layer of a dative subject, apparently obligatory for passive participles of ditransitives with dative subjects, but in general optional for other verbal forms with dative subjects. The features on the outer layer will then be visible to and trigger agreement on everything of which this dative is a subject, as illustrated in (16) below, where to reduce complexity, the generally optional *hafa verið* sequence is omitted. The structure uses the original 'OBJ2' label, which would now be replaced with 'OBJ$_\theta$', the semantic role to which these 'second objects' are restricted being Theme. The double-dotted line represents the feature-sharing between the complement subject and second object, which cashes out as feature sharing between the latter and the matrix subject thanks to the functional control represented by the solid line:

[10]In the instructions, '✓' was explicated as 'fully acceptable and natural, '?' a bit questionable ('acceptable, but perhaps somewhat unnatural'). '?*' as 'questionable', '*?' as 'worse, but not totally unacceptable', and '*' as bad.

(16)
$$\left[\begin{array}{ll} \text{SUBJ} & \left[\begin{array}{ll} \text{OBL} & \left[\begin{array}{ll} \text{PRED} & \text{`Pro'} \\ \text{PERS} & \text{III} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{MASC} \\ \text{CASE} & \text{DAT} \end{array}\right] \\ \text{NUM} & \text{PL} \\ \text{GEND} & \text{MASC} \\ \text{CASE} & \text{NOM} \end{array}\right] \\ \text{PRED} & \text{`Be(XCOMP)'} \\ \text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`Give(}\emptyset\text{, SUBJOBL, OBJ2)'} \\ \text{SUBJ} & [\quad] \\ \text{OBJ2} & \left[\begin{array}{ll} \text{PRED} & \text{`Horse'} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \text{MASC} \\ \text{CASE} & \text{NOM} \end{array}\right] \end{array}\right] \end{array}\right]$$

This provides a clever account of (14a), but not of (14b), since, given that the agreement features have been copied onto the shared complement subject and matrix object, they ought to be equally visible in both places. Andrews suggested that the acceptability of (b) was due to a performance effect caused by the greater distance between the agreement target and trigger, but it would be better to not have to resort to such explanations if possible.

The evident alternative is to have agreement with nominative objects (both OBJ and OBJ_θ) implemented by a second rule that applies if there is no suitable SUBJ agreement trigger. This is in general optional for non-passive verbs (but more optional or even dispreferred under circumstances investigated by Ussery (2017a) and many previous investigators, which we cannot pursue further here). This would be obligatory for passive participles, but a functional uncertainty expression to allow reference to a matrix dative subject would be optional.

The final problem is integration with the Kibort-Findlay Mapping Theory (KFMT). I will defer presentation of this problem until we have presented the proposed reanalysis of Quirky Case.

4 The Split Lexicon

The theoretical and empirical issues discussed in the previous section create difficulties for the two-level analysis; the idea we discuss here provides the infrastructure for the new one. This is the proposal for DBA Glue and the Split

Lexicon described in Andrews (2007, 2008). Its relevance is that it provides a rationale for distinguishing between case imposed by specific lexical items as opposed to structural rules, which requires less stipulation than other single-layer approaches, such as that of Sag et al. (1992).

Current Glue semantics performs compositional semantic interpretation by means of 'meaning constructors' in unitary lexical entries that contain all the information about a word and its meanings (Dalrymple, 2015). This works, but, as discussed in Andrews (2007, 2008), leads to awkwardness in some areas, such as with the numerous idiomatic verb-particle-preposition constructions in Germanic languages, and the lack of any explanation for the relatively consistent interpretation of grammatical features such as tense and number.[11] The original LFG analysis of Quirky Case was based on the idea that meaning was contributed by PRED-features. Therefore, the verb of such combinations, as well as that of ordinary idioms, would, would have PRED-features, while the other contributers would not, but would rather have other properties, such as CASE or FORM features. For example, a lexical entry for a combination such as *put up with* might get a (somewhat informal) lexical entry like this (assuming that idomatic prepositions introduce an attribute PFORM to avoid possibe FORM-feature clash):

(17) *put*: V, $(\uparrow\text{PRED}) = \text{`Put}_{up\ with}(\text{SUBJ, OBL[PFORM} =_c \text{WITH]}),$
$(\uparrow\text{DIR FORM}) =_c \text{UP}.$

This assumes that particles bear a grammatical function DIR, whose value can be a semantically meaningful directional, which would have its own PRED-features, but can also be a FORM-feature, which determines a morphological form with no independent meaning. This is workable, but it is awkward to provide so many forms with both PRED and FORM features, especially when the former are no longer playing a central role in semantics. And the original $(\uparrow(\downarrow\text{PCASE})) = \downarrow$ analysis of prepositional complements has not remained popular.

The proposal of the Split Lexicon is that in addition to the original LFG lexicon, in which the PRED-features functioned to a considerable extent as a substitute for a theory of semantics rather than a theory of semantics, there is also a Semantic Lexicon, where feature-values or constellations of feature-values are associated with meaning constructors. For *put up with*, we can represent a Semantic Lexicon Entry (SLE) as follows:

[11] The exceptions to this consistency, such as lexically determined grammatical gender and *pluralia tantum*, can be handled by allowing features to also be 'co-interpreted' with a lexical stem, similarly to idioms (Andrews, 2008, 8).

(18)
$$\begin{bmatrix} \text{SUBJ} & [\quad] \\ \text{PRED} & \text{‘Put’} \\ \text{DIR} & \begin{bmatrix}\text{PRED} & \text{‘Up’}\end{bmatrix} \\ \text{OBL} & \begin{bmatrix}\text{PRED} & \text{‘With’} \\ \text{OBJ} & [\quad]\end{bmatrix}\end{bmatrix} \Leftrightarrow e \multimap e \multimap t : \lambda yx.\mathit{Tolerate}(x,y)$$

The material to the right of the double arrow is the meaning-contribution, where positions in the f-structure are connected to argument-positions in the meaning-constructor. The meaning item ‘*Tolerate*’ should be seen as a placeholder for a substantive account of lexical meaning. The order of order of the ‘glue side’ to the left of the colon and ‘meaning-side’ to the right is swapped from the usual, to better fit this form of presentation.

What is relevant for us here is how the SLEs connects meaning-constructors to f-structures. The way this works is that the Morphological Lexicon (which, in terms of recent work on LFG morphology, such as Dalrymple (2015), should probably just be regarded as the Morphology) and the c-structure rules would produce an f-structure, and then the SLEs would apply, ‘checking off’ interpretable features (semantic case, but not structural case) subject to the constraint that each interpretable feature gets checked off once and only once.[12] Most SLEs check off only one feature, but (18) checks off three. The meaning constructors introduced by these SLEs are then assembled, subject to constraints of Glue semantics as presented in the literature, for example Dalrymple (2001) or Andrews (2010). If the verb is *walk* rather than *put*, there is no SLE that also checks off *with* and *up*, so something like *Jack walked up with Jill* requires the particle and preposition to choose one of their individual meanings, such as accompaniment and upward directionality.

There are various ways in which checking off could be implemented, but a natural way to do it would be to have interpretable feature-values having a pointer to the list of meaning-constructors introduced by the SLE (more than one is possible, as discussed in Dalrymple’s 2001 discussion of attributive adjectives). This can be ‘undefined’ for interpretable features that are not yet checked off, and a nil or list-terminating value for uninterpretable features. The result is that we obtain a natural distinction between the ‘Quirky’ cases that inhibit agreement, and the uninterpretable/structural ones that do not. We are now ready to present the analysis.

First, how do we implement regular case-marking, for which I think that LFG has never had a fully satisfactory account. The best so far is that of

[12]It would however be plausible to allow there to be ‘environmental’ features that must be present for an SLE to apply, but don’t get checked off. But it is not clear that this is necessary.

Nordlinger (1998), in which case-marked morphological forms introduce inside-out functional uncertainty (iofu) specifications for the environments they can (or must) appear in, but this account provides no explanation for why case *features* exist, even though they are clearly needed in many languages to explain phenomena of agreement. My suggestion here is that in addition to the 'semantic licensing' implemented by the checking-off mechanism, there is also 'structural licensing', implemented by iofu as presented by Nordlinger.[13]

Amongst the issues to be dealt with are a) how to deal with case on nonvert NPs (which creates problems for implementation in the PS rules as proposed by Neidle (1982)), b) how to make case-marking obligatory when it is (as is usually the case for the major standardized languages, but my impression from listening to field workers over the decades is that this is not always the case, especially for case-markers with a substantial degree of morphological autonomy). For the accusative in Icelandic, I provisionally propose this, superscript 0 on a feature meaning uninterpreted/structural:

(19) ACC: $((\text{OBJ}|\text{OBJ}_\theta \text{ CASE} \uparrow) \text{ SUBJ CASE}^0)$

This says that an ACC value of CASE is licensed if this occurs as the OBJ or OBJ_θ value of a structure that also has an uninterpreted SUBJ CASE-value. This will be nominative in a finite clause or anaphorically controlled að-infinitive, accusative in an ACI construction. Space limitations preclude saying more here about the justifications for various aspects of (19). The case features are assumed to be introduced into the f-structure by the morphology, including agreement, and (19) applies to any instance of the feature. Such syntactic licensing is then an alternative to licensing ('checking off') by a meaning-constructor.

A concomitant of (19) is that we have to analyse the nominative as another uninterpreted case-value, rather than the absence of any case-value. This value seems to behave as an 'elsewhere case', appearing where no other is licensed.[14] So we will need further provisions to require accusative to be present when its licensing condition is met, nominative otherwise, which can be done in various ways; the choice is not relevant here and so will not be discussed further.

5 A One-level Analysis

The basic generalization is that except with adjectival adjuncts as in (6a), certain control predicates (6b), and perhaps in a rather complex control construction (6c), both finite verbs and predicate adjectives and participles agree only

[13] And, on the basis of Butt & King (2003), there also appears to be dual licensing, for example, of cases on subjects expressing modality in Hindi and Urdu.

[14] In contrast to the behavior of the nominative in languages such as English, Modern Irish, and Ancient Greek, where the nominative seems associated with finiteness.

with uninterpreted case-values, that of their subject if this exists, otherwise, sometimes, with an object. Such agreement with subjects will be the main topic here, objects raising numerous questions of optionality and variation. Subject agreement can be restricted to non-Quirky NP triggers by including NOM^0 or ACC^0 in the agreement specifications, as illustrated in (20), where the full specifications will also include gender and number:[15]

(20) (↑SUBJ CASE) = NOM^0/ACC^0

Turning to the three cases in (6), we need to provide forms for all cases, not just nominative and accusative. Case (a), secondary predicates, can be managed by adding to an agreeing form (agreement target) a specification to the effect it must be a member of the ADJUNCTS grammatical function. This is easy to specify with an iofu constraint:

(21) (ADJUNCTS ∈ ↑)

Furthermore, when this specification is present, no requirement is imposed that the case of the agreement trigger be uninterpreted. I will not explore the kinds of notation that might be proposed to achieve this effect, since it clearly can be done with templates.

Case (b) is more challenging, because in current LFG, these would be treated as XCOMPs, just like 'Subject Raising' constructions (Bresnan et al., 2016, 289ff.). Therefore a simple extension of (21) to specify something like the original LFG ACOMP can't be used. But this requires these verbs to impose various kinds of category restrictions on their complements, because some of them take only a restricted range of possibilities, for reasons that are not entirely clear; explicable semantically to some degree, but not entirely:

(22) a. John grew unhappy.

b. *John grew a seasoned administrator. [must be transitive to be intelligible, therefore beyond current technology; contrast *became* instead of *grew*]

c. The tree grew into a fine provider of shade.

d. ?*John grew into enjoying syntax [seems off to me, construction not found from major dictionaries]

Given this need to impose properties on the complement, we can also impose one that permits agreement with NPs bearing interpreted case, which appears

[15]And should plausibly be reformulated to use something like AGR as in Alsina & Vigo (2017), but I will not pursue this here.

to be allowed only when the complement is restricted to not being verbal. The formulation is trivial.

The final case, (6c), involves examples in which Quirky accusative NP appears in ACI position of a passivized functional control verb (Andrews, 1990, 191). Some examples from the questionnaire referred to previously are (A.F.SG abbreviating ACC.F.SG):

(23) a. Þeir segja hana (vera) talið/talda vanta
they say her.ACC.F.SG (to have) been.SUP/ACC.F.SG to lack
peninga.
money
'They say that she is believed to lack money.'

b. Þeir segja strákana (vera) talda/*talið elska Svein.
they say her.A.F.SG (to have) been.ACC.F.SG/SUP to love Svein
'They say that she is believed to love Svein.'

The results were

(24)

	✓	?	?*	*?	*
a) SUP	1	3	0	1	2
a) AGR	5	2	0	0	0
b) AGR	5	1	1	0	0
b) SUP	0	0	0	0	7

It seems evident that (a) these sentences are not really very good (I have not managed to find such 'stacked functional control' constructions in web searches), and that agreement of the passive participle with a non-Quirky accusative overt object that is its f-structure subject is obligatory, but with a Quirky one, optional (indeed, agreement is better than non-agreement in this situation). Andrews (1990) suggested that the acceptability of agreement in (a) was due to a performance effect, due to the fact that the information that the accusative is Quirky is not provided until after the passive participle is produced, whereas, in the simpler and common examples, the Quirkiness of the subject is immediately evident, since it is sitting in a overt subject position.

This is plausible, but we should still look for ways of avoiding performance accounts of inconvenient data, and recent work on Icelandic and Faroese does reveal some threads to pull at. In particular, there is work indicating that agreement with Quirky NP is not actually as bad as originally thought. In Faroese, Jónsson (2009) showed that agreement with dative subjects was common enough to be reasonably regarded as grammatical, and Árnadóttir & Sigurðsson (2008) find some similar examples in Icelandic. In Andrews' 1982 original LFG analysis, extension of agreement to Quirky NPs would require either a reanalysis of the structures, or a complexification of the agreement conditions to allow access to the inner level, both somewhat complex (and it is unclear how to implement the former idea). On the present account, however, all that is necessary is to remove a restriction on the agreement specification, a more natural operation. The explanation for how the restriction got there in the first place would be diachronic: originally, the oblique subjects were not subjects.

6 Kibort-Findlay Mapping theory

We now show how to integrate the one-level analysis with the Kibort-Findlay Mapping Theory (KFMT). This is a version of lexical mapping theory that is fully explicit, formulable within the LFG formalism, and integrated with glue semantics. KFMT terminologically abandons the popular idea of 'argument structure', but replaces it with an elaboration of the 'semantic projection' of Glue semantics, which can perhaps be regarded as a kind of argument structure.[16] This is a projection from f-structure, which KFMT populates with attributes such as ARG_1, ARG_2 and more, which, in practice, partially reflect a classification of semantic roles in terms of their typical syntactic behavior.

ARG_1 is like the 'external argument' of GB/Minimalism, while ARG_2 is like the non-oblique 'internal argument' of GB/Minimalism. ARG_4 and below (with higher subscripts) are obliques, while ARG_3 is complicated, and will be discussed shortly below. KFMT also uses Davidsonian event semantics, with semantic projection attribute EV, so that verbs are fundamentally of type $ev{\rightarrow}t$.

A feature of current KFMT practice which I question here is that all (or perhaps most) arguments are added with templates that in effect attach the argument with its semantic role, in effecting converting a predicate of type τ into one of type $e{\rightarrow}\tau$ (that is adding another argument). This is workable for the commonly discussed semantic roles as Agent, Theme, Beneficiary, etc., but, as pointed out by an anonymous referee, is not required by the theory itself, and I think is rather questionable for the arguments of many verbs such as *predecease*, *outlive*, *survive* and *lack*. I don't think that anything goes wrong if we allow verbs to start out with some basic arguments, two at least, which is the maximum number that can take Quirky case.

This would give us an f-structure and semantic projection for a simple clause structure of an accusative subject and object verb such as *vanta* 'lack', where the assignments of ARG_2 and ARG_3 will be explained shortly. :

[16]Specifically, the 'lightweight' version of argument structure as proposed for example by Alsina (1996) or Andrews & Manning (1999) that imposes some classification and hierarchical ordering on the arguments, without digging into their semantics to any great extent.

(25)
$$\left[\begin{array}{ll} \text{SUBJ} & \left[\text{CASE} \quad \text{ACC}\right] \\ \text{PRED} & \text{`Lack'} \\ \text{OBJ}_{\theta} & \left[\text{CASE} \quad \text{ACC}\right] \end{array}\right] \qquad \left[\begin{array}{ll} \text{EV} & [\ \ ] \\ \text{ARG}_2 & [\ \ ] \\ \text{ARG}_3 & [\ \ ] \end{array}\right]$$

The SLE for *vanta* then ascribes the semantic roles to the two ARG-values, with an issue involving the CASE-values, as discussed further below

An essential component of KFMT is rules which equate ARG-values with GFs, which in effect apply optionally, via mechanisms not discussed here. The two relevant ones for this example are:

(26) a. $(\uparrow\{\text{SUBJ}|\text{OBJ}\})_{\sigma} = (\uparrow_{\sigma}\ \text{ARG}_2)$

b. $(\uparrow\text{OBJ}_{\theta})_{\sigma} = (\uparrow_{\sigma}\ \text{ARG}_3)$

Given the principles of the theory, the 'Lacker' argument needs to be ARG_2 in order to be subject (likewise for passivizable Quirky arguments such as the (dative) 'Helpee' of *hjálpa*), leaving the object OBJ_{θ} associated with ARG_3.

Using some notational shortcuts, we can now propose the following SLE (27) below for *vanta*. It accesses attributes of both f-structure (PRED) and s-structure (ARG and EV), consuming two type e arguments to produce a predicate over events (type $ev{\rightarrow}t$), using the standard convention that rightmost parentheses are omitted:

(27)
$$\left[\begin{array}{ll} \text{PRED} & \text{`Vanta'} \\ {}_{\sigma}\text{EV} & [\ \ ] \\ {}_{\sigma}\text{ARG}_2 & \left[{}_{\sigma}{}^{-1}\text{CASE} \quad \text{ACC}\right] \\ {}_{\sigma}\text{ARG}_3 & \left[{}_{\sigma}{}^{-1}\text{CASE} \quad \text{ACC}\right] \end{array}\right] \Leftrightarrow\ e{\rightarrow}e{\rightarrow}ev{\rightarrow}t : \lambda yxe.\mathit{Lack}(e,x,y)$$

The projections are a bit awkward-looking, but they could be eliminated with the aid of a 'coercion' convention similar to what most programming languages deploy when one mixes reals and integers in an arithmetic operation: functions are supplied automatically to make the types match in a useful way. In this case, when we see an s-structure attribute in an f-structure, we insert that σ projection, and, when we see an f-structure attribute in an f-structure, the σ^{-1} projection. Furthermore, it is necessary to interpret the inverse projections non-constructively, because even if the projections are functions, there is no guarantee that their inverses are, so the inverses need to be treated like iofu (a point which originated in some discussion with Mary Dalrymple and others).

We can now see the problem that the two-level analysis faces; not only would we need to specify the case, but somehow coordinate iofu specifications for a function such as OBL with rules such as (26), which is not necessarily impossible, but would still be a considerable nuisance, and is avoided by the present one-level analysis.

7 Conclusion

In conclusion, we see that Andrews' original 2-level analysis can be replaced with a 1-level analysis, where a major facilitating role is played by the proposal of the split lexicon, with semantics based on DBA of f-structure attributes rather than unitary lexical entries. This approach provides an independently motivated distinction between 'Quirky' and 'non-Quirky' case-values, which can control their differences in agreement behavior. A feature of this analysis is that 'Quirkiness' is not identified strictly with irregularity; there is plenty of evidence that the Quirky Datives are highly predictable, but from the meanings of lexical items rather than syntactic configurations. An interesting example of this that shows that more needs to be done in the application of KFMT to this material is the analysis of 'inversion' in Ussery (2017b), which shows that Quirky Case is fundamentally associated with semantic roles rather than s-structure attributes.

References

Alsina, Alex. 1996. *The role of argument structure in grammar*. Stanford, CA: CSLI Publications.

Alsina, Alex & Eugenio M. Vigo. 2017. Agreement: Interactions with case and raising. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of LFG17*, Stanford, CA: CSLI Publications.

Andrews, Avery D. 1982. The representation of case in Modern Icelandic. In Joan W. Bresnan (ed.), *The mental representation of grammatical relations*, 427–503. MIT Press.

Andrews, Avery D. 1990. Case-structures and control in Modern Icelandic. In *Modern Icelandic syntax*, 187–234. Academic Press.

Andrews, Avery D. 2007. Generating the input in OT-LFG. In J. Grimshaw, J. Maling, C. Manning & A. Zaenen (eds.), *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, 319–340. Stanford CA: CSLI Publications.

Andrews, Avery D. 2008. The role of PRED in LFG+glue. In Miriam Butt & Tracy Holloway King (eds.), *The proceedings LFG08*, 46–76. Stanford CA: CSLI Publications.

Andrews, Avery D. 2010. ‘Grammatical’ vs. ‘lexical’ meaning constructors for glue semantics. In Yvonee Treis & Rik De Busser (eds.), *Selected papers from the 2009 conference of the Australian Linguistic Society*, The Australian Linguistic Society.

Andrews, Avery D. & Christopher D. Manning. 1999. *Complex predicates and information spreading in LFG*. Stanford, CA: CSLI Publications.

Árnadóttir, Hlíf & Einar Freyr Sigurðsson. 2008. The glory on non-agreement: The rise of a new passive. Talk given at the Nordic Network for Intercultural Communication. URL: http://einarfreyr.wordpress.com/handrit/.

Asudeh, Ash. 2005. Control and resource sensitivity. *Journal of Linguistics* 41. 465–511.

Asudeh, Ash. 2012. *The logic of pronominal resumption*. Oxford University Press.

Barðdal, Jóhanna. 2011a. Lexical vs. structural case: A false dichotomy. *Morphology* 21. 619–654.

Barðdal, Jóhanna. 2011b. The rise of Dative Substitution in the history of Icelandic: A diachronic construction grammar account. *Lingua* 121. 60–79.

Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23. 757–807.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax, 2nd edition*. Blackwell.

Bresnan, Joan Wanda & Sam Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13. 181–254.

Butt, Miriam & Tracy Holloway King. 2003. Case systems: Beyond structural distinctions. In Ellen Brandner & Heike Zinsmeister (eds.), *New perspectives in case theory*, 53–87. Stanford, CA: CSLI Publications.

Dalrymple, Mary. 2001. *Lexical functional grammar*. Academic Press.

Dalrymple, Mary. 2015. Morphology in the LFG architecture. In *Proceedings of LFG15*, Stanford CA: CSLI Publications.

Findlay, Jamie. 2016. Mapping theory without argument structure. *Journal of Language Modelling* 4. 293–338.

Hale, Kenneth L. 1973. Person marking in Warlpiri. In Stephen Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 308–44. Holt, Rinehart and Winston.

Hale, Kenneth L. 1982. Some essential features of Warlpiri main clauses. In Stephen Swartz (ed.), *Papers in Warlpiri grammar: in memory of Lothar Jagst*, 217–315. Berrimah N.T.: Summer Institute of Linguistics: Australian Aboriginal Branch. *Work Papers of SIL-AAB*, Series A Volume 6.

Halvorsen, Per-Kristjan & Ronald M. Kaplan. 1995. Projections and semantic description in Lexical-Functional Grammar. In Mary Dalrymple, Ronald M.

Kaplan, John T. Maxwell & Annie Zaenen (eds.), *Formal issues in Lexical-Functional Grammar*, Stanford, CA: CSLI Publications.

Jónsson, Jóhannes Gísli. 2009. Covert nominatives and dative subjects in Faroese. *Nordlyd* 36. 142–164. URL: http://www.ub.uit.no/baser/septentrio/index.php/nordlyd/article/view/230.

Kaplan, R.M. & Joan Bresnan. 1982. Lexical-Functional Grammar: a formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Also in Dalrymple *et al.* (eds) 1995 *Formal Issues in Lexical-Functional Grammar*, pp. 29-130.

Laczkó, Tibor. 2012. On the (un)bearable lightness of being an LFG style copula in Hungarian. In Miram Butt & Tracy Holloway King (eds.), *Proceedings of LFG12*, 341–361. Stanford CA: CSLI Publications.

Neidle, Carol. 1982. Case agreement in Russian. In Joan W. Bresnan (ed.), *The mental representation of grammatical relations*, MIT Press.

Nordlinger, Rachel. 1998. *Constructive case*. Stanford CA: CSLI Publications.

Sag, Ivan, Lauri Karttunen & Jeffrey Goldberg. 1992. A lexical analysis of Icelandic case. In Ivan Sag & Anna Szabolcsi (eds.), *Lexical matters*, 301–318. CSLI Publications.

Sigurðsson, Halldór A. 2004. Non-nominative subjects: Facts and implications. In Peri Bhaskararao & K.V. Subbarao (eds.), *Non-nominative subjects*, vol. 2, 137–159.

Sigurðsson, Halldór A. & Anders Holmberg. 2008. Icelandic Dative Intervention: Person and Number are separate probes. URL: http://ling.auf.net/lingbuzz/@zdVxqlpSbDxoPPuM/jUFyAZYA.

Simpson, Jane & Joan Bresnan. 1983. Control and obviation in Warlpiri. *Natural Language and Linguistic Theory* 1. 49–64.

Simpson, Jane H. 1991. *Warlpiri morpho-syntax*. Kluwer Academic.

Thráinsson, Höskuldur. 2007. *The syntax of Icelandic*. Cambridge University Press.

Ussery, Charlon. 2017a. Agreement with nominative objects in Icelandic. In Höskuldur Thráinsson, Caroline Heycock, Hjalmar P. Petersen & Zakaris Svabo Hansen (eds.), *Syntactic variation in insular scandinavian*, 165–197. Amsterdam: Benjamins.

Ussery, Charlon. 2017b. Inversion as rightward dative-shift in Icelandic intransitives. In Sherry Hucklebridge & Max Nelson (eds.), *Proceedings of NELS 48*, North East Linguistic Society.

van Valin, Robert. 1991. Another look at Icelandic case marking and grammatical relations. *Natural Language and Linguistic Theory* 9. 145–194.

Number and Comitative-inclusory Constructions in Marori

I Wayan Arka
Australian National University and Udayana University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 48–68

http://csli-publications.stanford.edu/LFG/2019

Keywords: number, plural, comitative, inclusory, associative, morphosemantic, coordination, appositive

Arka, I Wayan. 2019. Number and Comitative-inclusory Constructions in Marori. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 48–68. Stanford, CA: CSLI Publications.

Abstract

The number system in Marori is morphosemantic, showing underspecified distributive exponence in number marking, which allows for the rich expression of nonsingular comitative-inclusory-associative meaning. The proposed LFG analysis consists of two parts: the decomposition of the number values ([+/-SG] and [+/-CUM]), and c-structure annotations with two template calls to capture the intricacies of the conjunctive and appositive properties involved in comitative constructions in Marori. It is demonstrated that the LFG analysis can straightforwardly capture the interaction of nominal and verbal number in expressing comitative meaning.

Keywords: number, plural, comitative, inclusory, associative, morphosemantic, coordination, appositive.

1 Introduction*

This paper discusses the interface of the morphology, syntax, and semantics of number in Marori with special reference to the comitative-inclusory constructions (Papuan, ISO:MOK, Merauke Indonesia), contributing to the existing typological and theoretical studies on this topic and related issues, including coordination (Corbett 1996, Stassen 2000, Sadler and Nordlinger 2010). The core meaning of comitative constructions, exemplified by (1) in English, is 'accompaniment', which involves an 'accompanee', a 'companion', and often a relator marking the comitative meaning (Stolz, Stroh, and Urdze 2006:2). Cross-linguistically, the comitative construction may also encode an inclusory meaning and this is often expressed by a pronominal form in Marori, such as example (2), in which a 'group' is part of the component meaning. In English, this 'plural/group' meaning can be added by the adverb *together*.

1	Obama	spends	his	vacation	together	with	his	daughters …
	(accompanee)				(group)	(relator)	(companion)	

The meaning components of the comitative-inclusory constructions are schematized in Figure 1. Languages differ in their manner of expression for these component meanings (cf. Lichtenberk [2000] for other Austronesian languages of the Pacific, such as Niuean, and Moravcsik [2003] for other languages, such as Hungarian).

* I gratefully acknowledge the support of ARC Discovery Grant (DP10100307) and ELDP MDP0336 (2016-17). For their stimulating discussion, I thank the audience at the LFG2019, in particular Miriam Butt and Rachel Nordlinger, as well as the anonymous reviewers whose comments have led to the improvement of the description and analysis presented in this paper. I also thank my Marori consultants (Pak Lukas, Pak Wiliam, and Mama Agustina) and my research assistants (Maxi Ndiken and Agustinus Mahuze) for their help throughout my research in Merauke.

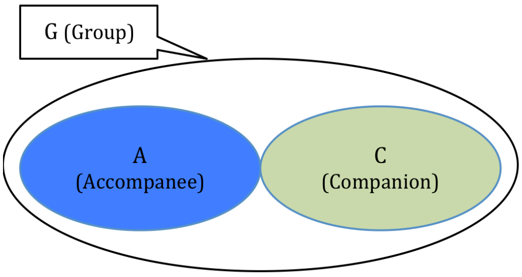


Figure 1: Conceptual elements of Comitatives

Comitative constructions in Marori are of special interest due to the fact that Marori has rich expressions demonstrating four different types in the same system (to be discussed and exemplified in section 2.2). Unlike the English example in (1), where the comitative unit prepositionally marked by *with* is an oblique not participating in the agreement with the verb, the equivalent structure in Marori shows that the comitative unit functions through coordination and, therefore, participates in the verbal agreement, as seen in (2). The singular subject ending *-du* is unacceptable because there are two participants involved in the event of sitting.[1] It is worthwhile noting that the free argument NPs in (2), such as *na*, have special discourse functions that are further discussed in section 3.1.

2	Na	kuye-den	/*kuye-du	Markus	fi
	1SG	sit.NPL-2DU.PRES	sit.NPL-1SG.PRES	Markus	COM
	'I am sitting with Markus.'				

In addition, distributed exponence in Marori number expression allows certain comitative meanings (such as the comitative-inclusory dual) to be expressed without inclusory dual pronouns (cf. example [9]).

The paper is structured as follows: after an overview of the morphosyntax and number system in Marori, the different types of comitative constructions in Marori are presented with examples. Then, the proposed LFG analysis is outlined with a demonstration of how the different comitative types and related issues can be accounted for. The conclusion is given in the final section.

[1] Abbreviations, alphabetically ordered: 1,2,3 (first, second and third person), A (Actor), AUX (auxiliary), CPLT (completive), AUX (auxiliary), COM (comitative), CUM(cumulative), DEIC (deictic), DU (dual), DUR (durative), F (feminine), FUT (future), HAB (habitual), LOC (locative), IRR (irrealis), MID (middle), NPL (nonplural), M (masculine), MP (macro present), NPL (Nonplural), NPST (near past), NSG (nonsingular), O (object), P (Patient), PART (particle), PL (plural), POSS (possessive), PRES (present), Q (question marker), REAL (realis), REL (relativiser), RPST (Remote Past), S (intransitive subject), STAT (stative), SG (singular), U (undergoer).

2 Marori comitative-inclusory constructions and their morphosyntax

2.1 Basic clausal morphosyntax and the Marori number system

Marori is a non-configurational verb-final language with the verb showing complex morphology that is inflected for tense, aspect, and mood (TAM). This TAM morphology is also used for number agreement, which is exemplified in (3) with the agreement suffix (*-ben* and *-ru*) indexing the Actor (A), or subject, while the prefix indexes the Undergoer (U). Marori also shows gender (U) agreement, however this is only applicable for the third person singular argument and is indicated by a vowel (/e/ 'masculine' vs. /o/ for feminine) on the verbal root, such as *-ife-* in (3a):

3 a. Na John=i ∅-ife-ben
1SG John=U 3SG.U-3SG.M.U.see-1NPL.NrPST
'I saw John.'

b. mbe=na kundo-ru
PART=1SG run-1SG.A.FUT
'I will run.'

The Marori number system is complex, showing an intricate nominal and verbal number interaction (see Arka [2012]). The nominal number system shows a three-way distinction (SG-DU-PL), but its formal morphological coding is underspecified as NSG or as NPL except for the first and second bound pronouns on the verb, as shown in (4).

4 Number coding system in Marori

		singular	dual	plural
Bound A pronouns on the verb	1 or 2	SG	DU	PL
	3A	NPL		PL
Bound U pronouns on the verb	3U	SG	NSG	
Free pronouns		SG	NSG	
Dem/spatial deictics		SG	NSG	
Derived event nominal		SG	NSG	
Common nouns*				

*with the exception of very few nouns

I argue that, unlike the English-type morphosyntactic number, the Marori number system is morphosemantic in nature (following Kibort [2010] and Corbett [2012]), allowing both the morphology and syntax to directly access and construct semantic number. The notable implication is that number 'agreement' in Marori differs significantly from that in English. Unlike in English, the free NP in Marori is optional, and its co-indexed bound pronominal on the verb is optionally referential.[2] In addition, the number of participants may also be encoded by the verbal number. Given these properties, it is not surprising that common nouns are not inflected for number (i.e. showing the so-called 'general number'). For example, the noun *purfam* in (5a) is interpreted as plural because it is understood as the sole participant of the intransitive plural verb 'stay'. However, in (5b) it is understood as singular due to the non-plural subject indexing *–fi*. Of course, the presence of the numerals *pnar*/*sokodu* also augments the plural/singular interpretation.

5 a.

sokodu	sour	ke	pnar	**purfam**	minggri-maf
one	house	LOC	three	person	stay.PL-3RPST

'in the house there were three people staying'
(FrogStory_Paskalis.004:00:00:42.962-00:00:49.852)

b.

mara	sokodu	**purfam**	kunonjo-fi	kier=ku
NEG	one	person	return-3NPL.RPST	village=LOC

'no one returned to the village'
(tete dan nene.077: 00:04:38.767-00:04:43.227)

Another important implication of Marori morphosemantic number, besides its underspecified morphological coding or distributed exponence of number coding, is the possibility of conveying constructed number. For example, dual meaning can be constructed by the NPL and NSG combination within a word or in the syntax between a free NSG pronoun and a co-indexed NPL suffix on the verb (see Arka [2012] for details).

[2] The bound pronoun on the verb is optionally referential because it may simply indicate agreement in cases where the reference is indefinite or generic, like in the following example:

Awo	*paya*	*kafra*
kangaroo	grass	3.eat.3PL.PRES

'Kangaroo eat grass.'

The Marori data point and the proposed analysis in this paper are in line with Simpson (1991) and Austin and Bresnan's (1996) analysis (contra to Jelinek's [1984] Pronominal Argument Hypothesis). Arguments with generic referents in Marori are always expressed with overt NPs. In this case, the NPs supply the semantic contents of the subject/object arguments, and the bound (zero) pronouns on the verb are simply agreement markers. In the LFG analysis, this is captured by having the PRED= 'pro' equation optional in the lexical entry of the subject/object affix. In most other cases, particularly when referents are definite (with no NPs present), and in cases involving comitative referents as discussed in this paper, the bound pronouns are referential.

Finally, the morphosemantic number in Marori allows the construction of even further number-related meaning: incompatible number values are not banned in syntax, but are, in fact, a resource exploited for a specific subtle number meaning. This is the case with plurality in relation to the comitative-inclusory construction, the focus of this paper, to which we now turn.

2.2 Types of comitative-inclusory constructions in Marori

As discussed earlier (cf. Figure 1), the core meaning of comitative constructions is the 'accompaniment' of A ('accompanee') and C ('companion'), with a reference to G ('group'). Cross-linguistically, elements C or G may be implied. Comitative constructions come with a cluster of related meanings, such as 'togetherness', 'proximity', and 'co-controlling of the events', which necessarily gives rise to semantic plurality. Their distinctive plural meaning is 'heterogenous plurality with reference to groups' (Daniel and Moravscik, 2005). This heterogeneous plurality is commonly observed in the plurality of pronouns. For instance, the first-person plural 'we' is a heterogeneous plural since it refers to a group of different people that includes the speaker and others, possibly with or without 'you' (inclusive/exclusive plural). This is different from a 'homogenous' or 'additive' plural as seen in common nouns, such as the plural noun *apples,* which refers to a group that is made up of the same kind of entities, each classifying as an 'apple'.

Cross-linguistically, the heterogeneous plural of accompaniment can be of different types, depending on how explicit or overt the accompaniment elements of A ('accompanee'), C ('companion'), and G ('group') are expressed and marked. Referring to Figure 1, there can be types such as 'comitative' (A + C), 'comitative-inclusory' (A + C + G), and 'inclusory-associative' (G + implicit/implied A/C). The English example in (1) represents the comitative type that is neither inclusory nor associative. Marori, however, does not have this type because the comitative construction in Marori requires NSG verbal agreement of the verb functioning as an inclusory pronoun. In what follows, different types of comitatives attested in Marori are discussed and exemplified, as summarised in Table 1.

Table 1: Comitative Constructions (CC) Types in Maori

	FREE NP ACOMMPANEE?	PRESENCE OF COMPANION MARKER *fi*	FREE INCL PRON?	INCL BOUND PRO ON THE VERB?
CC TYPE 1	Yes	Yes	No	Yes
CC TYPE 2	No	Yes	Yes	Yes
CC TYPE 3	No	Yes	No	Yes
CC TYPE 4	No	No	No	Yes

2.2.1 Comitative Constuction Type 1

Comitative Construction Type 1, the comitative-inclusory construction, is characterised by the overt presence of all elements (A, C, and G). This is exemplified by (6): the companion C (*Markus*) is a free NP flagged by the postposition *fi*. The accompanee (*na* '1SG') is also expressed by a free NP without flagging. The pronoun bound on the verb, -*den*, is inclusory, representing G. As seen previously in (2), forcing singular 'agreement' with -*du* '1SG.PRES' (i.e. treating it like the English example [1]) results in ungrammaticality. This suggests that the *fi*-flagged NP and the bare NP together participate in the verbal agreement. Therefore, unlike English *with*, *fi* in Marori functions like a coordinating conjunction.

6	Na	kuye-den	/*kuye-du	Markus	fi
	1SG	sit.NPL-1DU.PRES	sit.NPL-1SG.PRES	Markus	COM
	A	G		C	

'I am sitting with Markus.'

A and C can both be flagged by *fi*, however. This is shown by the data from a natural text in (7).[3]

7	mbe	tanamba	tok=i	eyew=nda-fi
	PART	now	frog=U	see=3.AUX-3NPL.RPST
				G
	Thomas=fi		koro=fi	
	Thomas=COM		dog=COM Thomas	
	A		C	

'Thomas and the dog were together looking for the frog'
(FrogStory_Paskalis.022: 00:02:03.630-00:02:07.320)

[3] Given that both group members ('Thomas' and 'dog') are flagged by *fi*, it is not immediately clear which is A or C. The focal member, which typically comes earlier (i.e. 'Thomas' in [7]) is analysed as A for pragmatic reasons.

2.3.1 Comitative Constuction Type 2

The Comitative Construction Type 2, exemplified by (8), is characterised by the presence of the group element G, possibly as free and bound pronouns, the companion C flagged by *fi*, and the implied accompanee A. The different pronouns expressing G in (8), *nie* '1NSG' and *-den* '1DU', refer to the same participant. They are appositive in nature, showing underspecified agreement and sharing the first person and non-singular values. The companion 'village chief' (C) shows up as a free NP flagged by *fi*. Contextually, it is understood as singular (i.e. there is only one village head). The accompanee (A), however, is implied and understood as '1SG'. This 'singular' interpretation is based on the following interpretation: the total number value of the set (i.e. a group number, which is 'dual') deducted by one (i.e. the 'singular' number value of C, 'the village head'). The notation [A] means that the comitative element A is implied.

8	nie	bab	desa	fi	uma-den	mukedu
	1NSG	uncle	village	COM	walk-1DU.PRES	middle
	G[A]	C			G[A]	

'I together with the village chief walk in the middle'
(BerburuPaskalisDkk18122011.134: 00:11:06.030_00:11:08.200)

Another example of the Comitative Construction Type 2 is given in (9). The inclusory G ('dual') is constructed by a NSG and NPL combination. Given that the companion C, *John,* is 'singular', then the implied number value of the accompanee A, 'you', can be worked out as 'singular'. That is, sentence (9) is only felicitous in the context where there is a single addressee, 'you', with *John* not being present at the moment of speaking.

9	Mba	kie	John=fi	korow=ku	war=na-ngga-fi?
	Q	2NSG	John=COM	hand=LOC	hold=RECIP-AUX.NPL-RPST
		G[A]	C		G[A]

'Did you (singular) and John marry with each other?'
(Lit. Did you and John hold hands with each other?'

2.3.2 Comitative Construction Type 3

Comitative Construction Type 3 is characterised by the presence of COM *fi* marking on C, and the absence of the free inclusory pronoun representing G. G only shows itself on the verb. In addition, there is no separate marking that is specific for A; it is only implied, as exemplified by (10) below:

10	**bab**	**desa**	**fi**	uma-**den**	mukedu
	uncle	village	COM	walk-1DU.PRES	middle
	C			G[A]	

‘I together with the village chief walk in the middle’

Example (11) shows a Type 3 with a comitative-associative meaning, where both C and (implied) A are third person participants: John is accompanied by an implied nonspecific associate (reading [ii]). Note that the inclusory NPL on the verb elicits a singular interpretation of the implied associate A (i.e. the unacceptability of a plural interpretation, reading [iii]). Sentence (11) cannot have a non-comitative interpretation either (i.e. the unacceptability of reading [i]). If a non-comitative reading is intended, then the COM marker *fi* must not be used, as shown in (12), reading (i):

11	John	fi	kier=i	ki=ngge-∅-f
	John	COM	village=U	leave=3SG.M.AUX-3NPL-NPst.PF
	C			G[A]

i. * ‘John left the village.’
ii. ‘John and his associate left the village.’/
‘With John included, they (NPL/two) left the village.’
iii. * ‘John and his associates left the village.’

12	John	kier=i	ki=ngge-∅-f
	John	village=U	leave=3SG.M.AUX-3NPL-NPst.PF

i. ‘John left the village.’
ii. * ‘John and his associate left the village.’
iii. * ‘John and his associates left the village.’

2.3.3 Comitative Construction Type 4

Comitative Construction Type 4 is characterised by the exploitation of incompatible number values in the agreement system in Marori; the plural number of the verb functions inclusorily to express G when it co-occurs with a singular free argument NP. This is exemplified in (13): *John*, the singular subject NP, is made compatible with the plural intransitive verb (appearing with the pluractional *-fre*), giving rise to an inclusory-associative reading and implying that there are other companion participants, as shown by reading (iii). Note that reading (i) (non-comitative) is not acceptable. Reading (ii) (i.e. the comitative-inclusory with a singular associate) is also not acceptable because the total number value of the group (G) (including 'John') is plural, and plural in Marori is 'more than three' (Arka 2011, 2012):

13 John kier=i ki=ngge-fre-fi
John villlage=U leave=AUX-PL-3RPst
A G[C]
i. * 'John left the village.'
ii. * 'John and his associate left the village.'
iii. 'John and his associates left the village.'

As shown in (14), flagging the free NP with *fi*, making it CC Type 3, is also acceptable with the same logical meaning. The only subtle difference seems to be that (14) has the pragmatic focus of *John* as the companion (C) rather than the accompanee (A).

14 John fi kier=i ki=ngge-fre-fi
John COM villlage=U leave=AUX-PL-3RPst
C G[A]
i. * 'John left the village.'
ii. * 'John and his associate left the village.'
iii. 'With John, his associates left the village.'

Sentence (15) is another example showing that incompatible number values trigger a comitative-inclusory interpretation. In this example, the values of the PERS feature are also incompatible: '3' of Markus vs. '1' of *-den*. It should be noted that while the absence of the COM marker *fi* is acceptable, sentence (15) is preferred with the presence of *fi*:

15 Markus (fi) kuye-**den**
Markus COM sit.NPL.1DU.PRES
'With Markus, I am sitting.' or 'Markus and I are sitting together.'

The comitative meaning of 'togetherness' also applies to inanimate arguments, which are typically associated with objects. Consider Type 4 in (16)[4], exploiting the verbal number resource in Marori: the verb root *kei* is in plural form, requiring a plural object argument, but the object NP is singular, as indicated by the adjectival singular noun *anep* 'big.SG' and the numeral *sokodu* 'one'. The sentence means that the object is a collection of entities, with one of them being a big coconut; the things carried could be other coconuts of different sizes and/or other entities.

[4] It remains a question for future research whether comitative constructions with inanimate arguments are possible under other CC types.

16	fis	[anep	poyo=i	sokodu]	[kei-ben]
	yesterday	big.SG	coconut=U	one	bring.NSG-1NPL.NrPST

‘one big coconut, I/we (two) carried it together with other coconuts/things’

3 LFG analysis

The components of the proposed LFG analysis consist of two parts: fine-grained referential (number and person) features in lexical entries and phrase structure annotations capturing the different status of free NPs and their (dis)agreeing pronouns on the verb. All comitative types in Marori show properties associated with coordinating/asyndetic structures. Following earlier work on coordination and nominal juxtaposition in LFG (Sadler and Nordlinger 2010, Dalrymple and Kaplan 2000, Dalrymple 2001), I will demonstrate that the complexity of comitative-inclusory constructions in Marori is treated as having set-valued grammatical functions.

3.1 Number (NUM) and person (PERS) features

Given the morphologically rich number system in Marori, especially its underspecification and distributive exponence in number coding that allows a constructed number (Arka 2012), I adopt an analysis whereby number values of ‘singular’, ‘dual’, and ‘plural’ are decomposed into features, as shown in (17). The features +/−SING and +/−CUM refer to ‘singular’ vs. ‘nonsingular’ and ‘cumulative’ vs. ‘noncumulative’, respectively (Arka and Dalrymple 2013). CUM(ulative) is a property of cumulative reference. Only the plural (which, in Marori, is ‘three or more’) is [+CUM], whereas ‘singular’ or ‘dual’ is [−CUM].

17 Decomposition of number values

‘Singular’	‘Dual’	‘Plural’
+SG	−SING	−SING
−CUM	−CUM	+CUM

I also follow Dalrymple and Kaplan (2000) in representing PERS values in terms of a set of S ‘1st’, H ‘2nd’, and { } ‘3rd’; hence, the following represents the first person inclusive/exclusive NSG:

18 a. {S} U {H} = {S, H} ‘1st INC’

b. {S} U { } = {S} ‘1st EXC’

The lexical entry for the dual/plural pronoun contains the referential features signifying a set of non-homogenous referents. For example, the entry for the first person dual bound pronoun on the verb, *-den*, as seen in sentence (15), can be represented as (19). Its information is then passed up to become the value of SUBJ during the process of word formation (briefly discussed in the next subsection).

19 *-den*
(($\uparrow$PRED)= ‘pro’)
($\uparrow$INDEX NUM) = [-SG,-CUM]
($\uparrow$INDEX PERS)= {S, H}
($\uparrow$TNS) = PRES

3.2 Phrase structure analysis and GF annotation

The representation in (20a) shows the main c-structure properties of the Marori clausal structure. Further, it demonstrates that the language is verb-final, and that a sentence may be analysed as having an extended clause structure with a sentence-initial free XP, bearing a discourse function (DF) such as TOP or FOC (see Arka [2017] for the details of the information structure in Marori). In between, there can be other XPs (freely ordered) and a predicative nominal (N:PRED), which typically immediately precedes the inflected light or auxiliary (AUX) verb.

The bound pronominals on the verb are arguments bearing SUBJ and OBJ functions. A unit marked by the COM marker *fi* is a PP (20c) in which *fi* is a postposition whose lexical entry is shown in (20d). The free XP can be a NP or a PP (20b), carrying different template calls (represented by @), whose details are shown in (20f-g): two (either @CNJT or @APPOS) for the NP and one (@CNJT) for the PP.

20 a. CP (Extended Clause)

XP
($\uparrow$DF) =$\downarrow$

S (Core Clause)

XP* N:PRED [PREF-AUX.V-SUFF]V
($\uparrow$OBJ)=$\downarrow$ ($\uparrow$SUBJ)=$\downarrow$

b. XP = {NP: {@CNJT | @APPOS} |
PP: @ CNJT}

c. PP $\rightarrow$ NP P

d. *fi* P ($\uparrow$PCASE) = COM

e. DF = {TOP, FOC}
GF = {SUBJ, OBJ, OBL, ADJUNCT }

f. CNJT:	g. APPOS:
$\downarrow \in (\uparrow \text{GF})$	$\downarrow \in (\uparrow \text{GF})$
($\downarrow$INDEX PERS) ⊆ ($\uparrow$GF INDEX PERS)	($\downarrow$INDEX) = ($\uparrow$GF INDEX)
($\downarrow$INDEX GEND) ⊆ ($\uparrow$GF INDEX GEND)	

The templates @CNJT and @APPOS in (20f-g) follow Sadler and Nordlinger's (2010) formulations, which capture conjunctive and appositional specifications, respectively. The only difference is the notation $\downarrow \in (\uparrow \text{GF})$, which means that the information (regarding the XP) is part of the set-valued GFs. This captures an important empirical point in Marori, specifically that comitative constructions exhibit properties associated with coordinating/asyndetic structures.

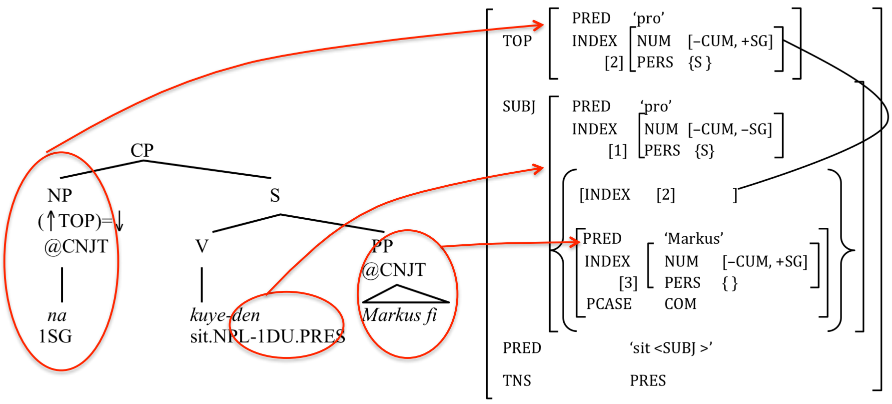
Importantly, the template @CNJT (20f) expresses that the person (PERS), number (NUM) and gender (GEND) features are part of an INDEX, which is non-distributive (i.e. resolving), whereas the constraint of @APPOS makes INDEX values distributive (i.e. identical to the mother node). @CNJT creates the effect of a unification in which a {S, H}(1st person) unit combines with a {H}(2nd person) unit to be resolved in becoming {S, H}(1st person) (see Dalrymple and Kaplan [2000] for details). In contrast, the @APPOS specification results in the spread of the INDEX values to the mother node. The availability of the two options (@CNJT or @APPOS) for an NP argument leads to the (im)possibility of a particular number and comitative interpretation, as will be discussed in the following subsections.

3.3 Demonstration and discussion

This section demonstrates how the proposed analysis works to account for each CC type and discusses relevant issues. We start with CC Type 1, represented by sentence (6) and repeated as (21). The NP TOP *na* carrying [+SG] must also be identified to bear a GF, which is SUBJ in this case. Hence, it can only have the @CNJT option because the @APPOS option would make its NUM value clash with the dual SUBJ -*den*, which carries [–SG]. The @CNJT specifications result in the referent (INDEX) of *na* being understood as a subset of SUBJ. Likewise, the f-str contents of the PP *Markus fi* result in a subset of the SUBJ value, as expected, yielding the right comitative interpretation.

21 Na kuye-den Markus fi
1SG sit.NPL-1DU.PRES Markus COM
‘I am sitting with Markus.’

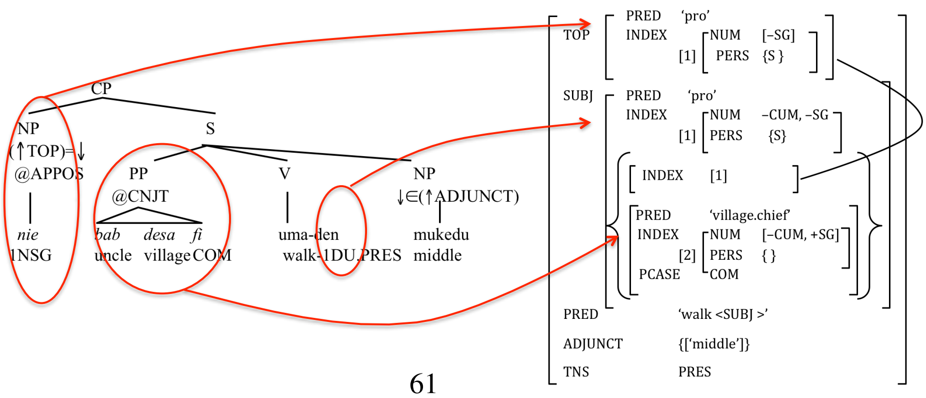
C-str and f-str of (21):



CC Type 2 is illustrated by sentence (22), which demonstrates an appositive relation of inclusory pronouns in the sentence. Its c-str and f-str representations are given in (23). The TOP NP *nie* can take the @APPOS option, and its index is also identified as the SUBJ (tag [1]) because its number value carries [–SG], and, therefore, it can unify with that of the dual SUBJ -*den* coming from the verb (carrying [–CUM, –SG]). Note that *nie* is underspecified, carrying no CUM feature in its entry, as it is usable for ‘dual and plural’. The comitative PP ‘the village head’, as expected, can only have its contents as a subset of the value of SUBJ.

22 a. nie bab desa fi uma-den mukedu (= 10)
1NSG uncle village COM walk-1DU.PRES middle
‘the village chief and I walk in the middle’, or
‘we (two) including the village chief walk in the middle’

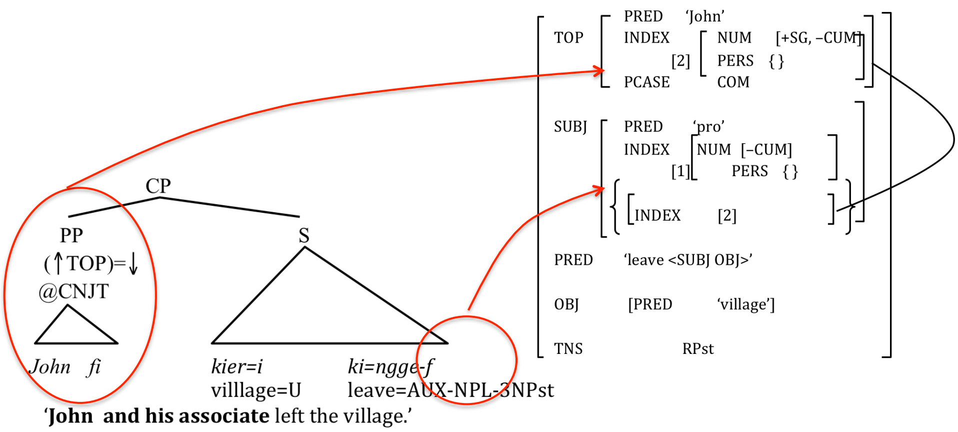
23 C-str and f-str of (22).



CC Type 3 can be accounted for straightforwardly. Consider the following example in (24), which results in a dual comitative-associative meaning. The c-str and f-str are shown in (25). The COM PP *John fi* can only have the conjunctive specifications, and its INDEX (tag [2]) is a subset of the SUBJ value. Given that its NUM value is [+SG] and that the inclusory SUBJ is [–CUM], then the plural reading (iii) is excluded because plural is [+CUM]. Further, reading (i) is also excluded as *fi* explicitly marks *John* as a companion member in a group/set together with another participant (see the conception in Figure 1):

24 John fi kier=i ki=ngge-∅-f (=11)
John COM village=U leave=3.AUX-3NPL-NPst.PF
i. * 'John left the village.'
ii. 'John and his associate left the village'/
'With John included, they (NPL/two) left the village.
iii. * 'John and his associates left the village.'

25 C-str and f-str of (24):

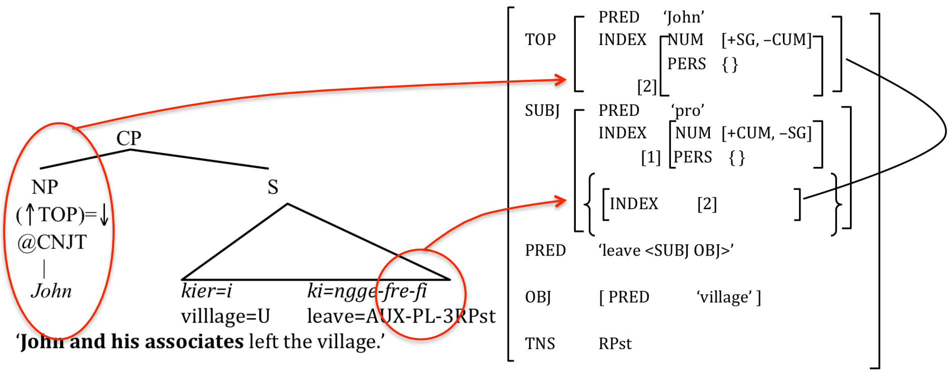


It should be noted that the inclusory-accompanee/associate meaning within a group is only implicit in the f-str representation in (25). It is clearly part of the constructed inclusory dual number in Marori. To be explicit, it is perhaps necessary to enrich the feature structure of (17) with [+/– GRP(group)] (cf. Jones [2015], building on work by Sadler [2011], Arka [2011, 2012] and Nordlinger [2012]). In this revised feature structure analysis, 'dual' would be represented as [–SG, –CUM, +GRP]. A discussion on the implications of adding [+/– GRP] to the currently adopted feature system of number is beyond the scope of this paper.

Recall that CC Type 4 is characterised by the absence of the *fi* flagging and the exploitation of incompatible number values. This type can also be nicely captured in the proposed analysis through the use of the @CNJT annotation. Consider (26), whose c-str and f-str are represented in (27). Note that while the NP can have either @CNJT or @APPOS, only @CNJT is applicable because @APPOS would result in a clash with the number value of *John* ([+SG]) and that of SUBJ, which carries [–SG]. As seen in (27), the @CNJT option results in *John* being interpreted as a member of the plural set ([+CUM, –SG]). It correctly captures the plural associative meaning that John's companion is not singular in number.

26 John kier=i ki=ngge-fre-fi
John village=U leave=AUX-PL-3RPst
i. * 'John left the village.'
ii. * 'John and his associate left the village.'
iii. 'John and his associates left the village.'

27 C-str and f-str of sentence (26).

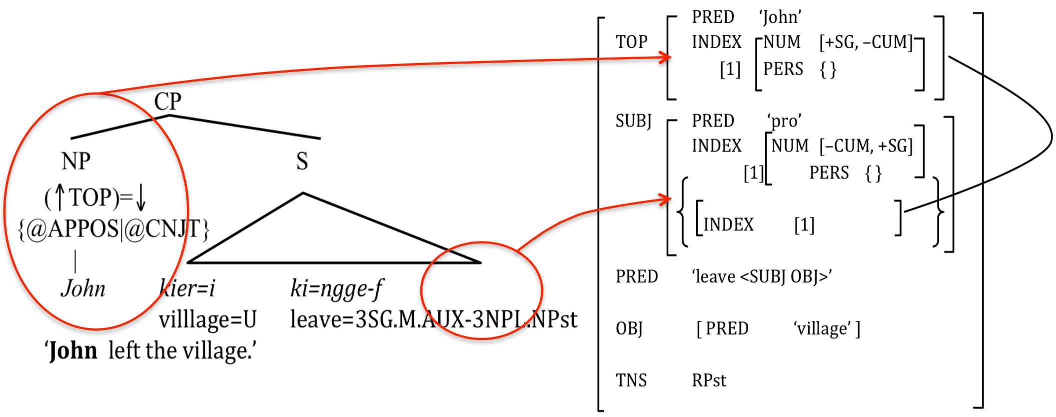


There is an issue with the non-comitative reading of the structure of the type exemplified in (28) that merits further discussion. The c-str and the f-str of the non-comitative reading, due to the application of the @APPOS associated with the NP (*John*), are shown in (29). This is straightforward, with the singular value ([+SG]) spreading and giving rise to reading (ii).

Reading (iii) is expected to be unacceptable because of the NPL value of the subject. However, reading (ii), which is expected to be acceptable, turns out to be unacceptable as well. Further, reading (ii) is expected to be acceptable if the @CNJT option is applied. Recall that this option is available for an NP, and it is indeed applicable, as seen in (26).

28 John kier=i ki=ngge-f
John village=U leave=3SG.M.AUX-3NPL.NPst.PF
i. ‘John left the village.’
ii. ?* ‘John and his associate left the village.
iii. * ‘John and his associates left the village.’

29 C-str and f-str of sentence (28).



A close examination of the case in (28) reveals that the free NP and the SUBJ carry harmonious referential features: *John* comes with [PERS {}] and [NUM +SG, –CUM], and the verb comes with [NUM –CUM]. This is important, and I propose the presence of a preference constraint for agreement (in line with the elsewhere or Paninian rule). That is, agreement with harmonious features in Marori grammar is unmarked for which feature spreading (captured by @APPOS) applies, whereas the @CNJT application is a marked ‘disharmonious’ type of agreement. As we have seen, the overt marking of @CNJT is performed by *fi* in Marori. In other words, when the structure is unmarked and contains harmonious agreement features, then @APPOS applies. Otherwise, if it is flagged by *fi*, then @CNJT must apply, thus blocking @APPOS.

However, it should also be noted that @CNJT can apply without *fi* when the supposedly agreeing arguments carry disharmonious referential features. Evidence for this comes from data points shown in sentence (15), repeated here as (30):

30 Markus (fi) kuye-**den**
Markus COM sit.NPL.1DU.PRES
‘With Markus, I am sitting.’ or ‘Markus and I are sitting together.’

As seen, the dependent argument *Markus* ([–CUM, +SG]) and the SUBJ pronoun *-den* ([–CUM, –SG]) are not harmonious in terms of the NUM feature. This 'disharmonious' agreement licenses the application of @CNJT to the dependent. Hence, the flagging by *fi* is optional, as shown by placing *fi* within brackets. That is, even if the NP *Markus* is without *fi*, it receives the conjunctive specifications, resulting in the comitative reading.

Finally, the comitative construction in Marori which makes use of verbal number can also be captured straightforwardly in LFG. The relevant example (16) is repeated here as (31). The verb 'bring/carry' shows a suppletive root alternation (*kei* ~ *ndV*), encoding the participant number of the undergoer, in combination with the pluractional suffix *–rV,* which encodes event plurality and may also encode the plurality of the actor participant.[5] The two give rise to the stem forms shown in (32).

31	fis	[anep	poyo=i	sokodu]	[kei-ben]
	yesterday	big.SG	coconut=U	one	bring.NSG-1NPL.NrPST

'one big coconut, I carried it together with (an)other coconut(s)/thing(s)'

32

	ROOT ALTERNATION:	
PLURACTIONAL SUFFIX:	'bring.SG.U'	'bringNSG.U'
NPL	(i) *ndV-∅*	(ii) *kei-∅*
PL	(iii) *ndV-rV*	(iv) *kei-rV*

The different participant number requirements can be specified in the lexical entries of the roots, as shown in (33).

33	a. *ndV*	b. *kei*
	(↑PRED)= 'bring<SUBJ, OBJ>'	(↑PRED)= 'bring<SUBJ, OBJ>'
	(↑OBJ INDEX NUM)= +SG	(↑OBJ INDEX NUM)= –SG

As discussed earlier, all dependent argument NPs can have an alternative option of either @APPOS or @CNJT, thus the sentence (31) will have disharmonious referential features in its INDEX path. The @APPOS option would lead to a clash (i.e. unacceptable). The @CNJT option provides a way out, but it gives rise to a comitative heterogeneous plural reading. The f-structure is shown in (34). It should be noted that the nonsingular group reading of OBJ is underspecified; it is typically

[5] The notation *V*, in *ndV*- and *rV*, refers to a 'vowel' whose specific value is determined by vowel harmony, and it encodes gender and number inflection; see Arka (2011, in press) for details.

heterogeneous ‘plural’ rather than ‘dual’, but its interpretation is a matter of context.

34
$$
\left[\begin{array}{ll}
\text{PRED} & \text{‘bring<SUBJ OBJ>’} \\
\text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{‘pro’}\\ \text{INDEX} & \text{1NPL}\end{array}\right] \\
\text{OBJ} & \left[\begin{array}{l}\text{INDEX} \quad [\text{NUM} \quad -\text{SG}] \\ \left\{\left[\begin{array}{ll}\text{PRED} & \text{‘coconut’}\\ \text{INDEX} & [\text{NUM} +\text{SG}]\\ \text{SPEC} & \text{‘one’}\\ \text{ADJUNCT} & \{\text{‘big.SG’}\}\end{array}\right]\right\}\end{array}\right] \\
\text{ADJUNCT} & \{\text{‘yesterday’}\}
\end{array}\right]
$$

4 Conclusion

This paper has discussed the number system in Marori with special reference to comitative constructions. The data points demonstrate how incompatible referential (NUM/PERS) features in Marori are acceptable and give rise to a comitative reading, strongly suggesting that agreement in this language is quite different from those in Indo-European languages, such as English. The Marori number system is morphosemantic in nature and not morphosyntactic as in English. It allows the exploitation of referential NUM/PERS resources to express rich nonsingular (i.e. ‘dual’ and ‘plural’) meanings, including those identified as ‘comitative-inclusory-associative’. This is particularly possible due to the distributive underspecified coding of number, which allows the so-called ‘constructed’ number in the Marori language (a salient feature that Marori shares with its neighbouring languages in southern New Guinea [Evans et al. 2017], such as Ngolmpu [Carroll 2016]).

It has been demonstrated in this paper that LFG is well equipped with the machinery to capture the intricacies of the Marori number system. The comitative constructions in Marori can be straightforwardly analysed in LFG by means of template call annotations on c-str. The template calls (@CNJT and @APPOS) capture the essence of their conjunctive and/or appositive properties. Further, the analysis accurately captures the intricacies of how nominal and verbal number interact in constructing comitative-inclusory meanings in this language.

More research is needed, however, to uncover the full extent of the interaction of nominal and verbal number in Marori within a typological context, as well as its theoretical significance. Marori, for example, has 'paucal' referring to a small group of entities in contrast to a big one (not discussed in this paper; see Arka [in press]). The full analysis of 'paucal' in Marori and its interaction with comitative meaning is beyond the scope of the present paper. It requires refinement of the proposed number feature decomposition given in (17). Such future research will ideally incorporate evidence from textual/corpus-based evidence.

5 References

Arka, I Wayan. 2011. "Constructive number systems in Marori and beyond." In *The proceedings of the LFG'11 conference*, edited by Miriam Butt and T. Holloway King, 5-25. Stanford: CSLI Publications http://csli-publications.stanford.edu/

Arka, I Wayan. 2012. "Verbal number, argument number and plural events in Marori." In *Proceedings of the LFG'12 conference*, edited by Miriam Butt and Tracy H. King, 23-43. Stanford: CSLI Publications http://csli-publications.stanford.edu/

Arka, I Wayan. 2017. "Information structure and its morphosyntactic resources in Marori." In *Proceedings of the the third International workshop on information structure of the Austronesian languages*, edited by Atsuko Utsumi and Asako Shiohara, 73-89. Tokyo: ILCAA, Electronic Publication, TUFS; https://publication.aa-ken.jp/.

Arka, I Wayan. In press. "Number in Marori, a Papuan language of Indonesia." In *The Oxford Handbook of Number in Language*, edited by Patricia Cabredo Hofherr and Jenny Doetjes. Oxford: Oxford University Press

Arka, I Wayan, and Mary Dalrymple. 2013. Expression of number: From morphosyntax to semantics. Canberra, unpublished ms.

Austin, Peter and Joan Bresnan. 1996. "Nonconfigurationality in Australian Aboriginal languages." *Natural Language and Linguistic Theory,* 14:215-68.

Carroll, Matthew. 2016. *The Ngkolmpu Language with special reference to distributed exponence*, Linguistics Department, Australian National University, Canberra, Australia.

Corbett, Greville G. 1996. "Associative forms in a typology of number systems: evidence from Yup'ik." *Journal of Linguistics,* 31:1-17.

Corbett, Greville G. 2012. *Features*. Cambridge, UK: Cambridge University Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, *Syntax and semantics*. San Diego: Academic Press.

Dalrymple, Mary, and Ronald M. Kaplan. 2000. "Feature indeterminacy and feature resolution." *Language,* 76 (4):759-798.

Evans, N, W. Arka, M. Carroll, C. Döhler, E. Kashima, E. Mittag, K. Quinn, J. Siegel, P. Tama, and C. van Tongeren. 2017. "The Languages of Southern New Guinea." In *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, edited by B. Palmer, 641-774. Berlin: De Gruyer Mouton.

Jelinek, Eloise. 1984. " Empty Categories and Non-Configurational Languages." *NLLT,* 2:39-76

Jones, Stephen M. 2015. *Number in Meryam Mir*. M.Phil. dissertation, University of Oxford.

Kibort, Anna. 2010. "Towards a typology of grammatical features." In *Features: Perspectives on a Key Notion in Linguistics*, edited by Anna Kibort and Greville G. Corbett, 64–106. Oxford: Oxford University Press.

Lichtenberk, Frantisek. 2000. "Inclusory pronominals." *Oceanic Linguistics,* 39 (1):1-32.

Moravcsik, Edith. 2003. "A semantic analysis of associative plurals." *Studies in Language*, 27 (3):469-503.

Nordlinger, Rachel. 2012. "Number marking in the Daly languages (Australia)." In *Proceedings of the LFG'12 Conference*, edited by M Butt and T. H King, 422-439. Stanford: CSLI Publications http://csli-publications.stanford.edu/

Sadler, Louisa. 2011. "Indeterminacy, complex features and underspecification." *Morphology,* 21 (2):379-417.

Sadler, Louisa, and Rachel Nordlinger. 2010. "Nominal juxtaposition in Australian languages: An LFG analysis." *Journal of Linguistics*, 46 (02):415-452.

Simpson, Jane. 1991. *Warlpiri-Morpho Syntax*. Dordrecht: Kluwer.

Stassen, Leon. 2000. "AND-languages and WITH-languages." *Linguistic Typology,* 4 (1):1-54.

Stolz, Thomas, Cornelia Stroh, and Aina Urdze. 2006. *On comitatives and related categories*. Berlin: *Mouton de Gruyter*

The Syntactic Encoding of Information Structure in the History of Icelandic

Hannah Booth
Ghent University and University of Konstanz

Christin Schätzle
University of Konstanz

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 69–89

http://csli-publications.stanford.edu/LFG/2019

Keywords: syntactic change, icelandic, information structure, verb position

Booth, Hannah, & Schätzle, Christin. 2019. The Syntactic Encoding of Information Structure in the History of Icelandic. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 69–89. Stanford, CA: CSLI Publications.

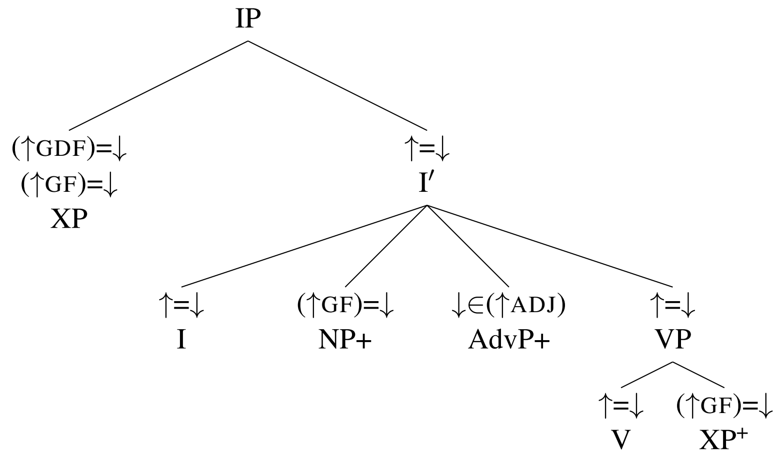
Abstract

In this paper, we present a revised LFG account for Icelandic clause structure, factoring in new historical data from IcePaHC (Wallenberg et al., 2011). This builds on previous work by Sells (2001, 2005) and Booth et al. (2017), focusing more closely on the syntactic encoding of information structure. Based on findings from a series of corpus-based investigations, we argue that the functional category I was already obligatory in Old Icelandic, accounting for both V1 and V2 orders and the absence of V3/V-later orders. Moreover, we show that the basic c-structure skeleton persists throughout the diachrony; what changes is the way in which information structure is syntactically encoded, i.e. the association between c- and i-structure. Topics increasingly target SpecIP, which allows the finite verb in I to serve as a boundary between topic and comment. This goes hand in hand with certain discourse adverbs losing their function as a discourse partitioner in the midfield and ties in with other changes shown for Icelandic (Booth et al., 2017).

1 Introduction

The clause structure of modern Icelandic has attracted a good deal of attention in generative syntax. Within LFG, Sells (2001, 2005) gives the overall structural possibilities for matrix clauses as (1).

(1)



The clause is rooted in IP, headed by a functional head I which is associated with finiteness. SpecIP is an information-structurally privileged position which can host any GF (grammatical function) with a GDF (grammaticalised discourse function; Bresnan et al., 2016). Within I′ there is a flat 'midfield' area, bounded by I and the VP, in which any GF not associated with a discourse function can occur, as well as

†We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding within the project "Evaluation Metrics for Visual Analytics in Linguistics" – Project-ID 251654672 – TRR 161. We would also like to thank Miriam Butt, Kersti Börjars, and the audience of the LFG19 conference at ANU Canberra for their valuable feedback on this work.

any number of ADJ(unct)s (see also Börjars et al., 2003 on Swedish). According to Sells, the linear order of these midfield elements is somewhat free but subject to OT constraints (see also Börjars et al., 2003).

A revised LFG account of Icelandic clause structure is proposed by Booth et al. (2017) who focus on the left periphery. This proposal is shown in (2).

(2)

$$
\begin{array}{cc}
 & \text{IP} \\
\left\{\begin{array}{l} (\uparrow \text{TOPIC}) = \downarrow \\ (\uparrow \{\text{COMP}|\text{XCOMP}\}^{*}\ \text{GF}) = \downarrow \\ \hline (\uparrow \text{EXPLETIVE}) =_{c} + \\ \neg(\uparrow \text{TOPIC}) \end{array}\right\} & \text{I}' \\
\text{XP} & \text{I} \quad \ldots
\end{array}
$$

Like Sells, this revised account assumes that matrix clauses are IPs with the finite verb in I, and that SpecIP is associated with a GDF (specifically TOPIC). Alternatively, SpecIP can host an expletive, provided the clause lacks a topic. Unlike Sells, Booth et al. (2017) assume that the expletive is a topic position placeholder, not a subject (following Thráinsson, 1979, Maling, 1988, Sigurðsson, 2007). A third option is that SpecIP can be unoccupied, rendering topicless V1 (verb-first) sentences.

Icelandic has a rich historical attestation which dates back to the 12th century. Data from earlier stages of the language challenges the accounts by Sells (2001, 2005) and Booth et al. (2017) in a number of ways.[1] Firstly, topics do not exclusively occur in the clause-initial prefinite position. V2 sentences with a clause-initial non-topic, and an immediately postfinite topic are common, e.g. (3).[2]

(3) a. En fullt var **skipið**. (AdjP-V-TOPIC)
but full.NOM be.PST ship.NOM.DEF
'But the ship was full.' (1210, Jartein.779)

b. Þá hafði **hann** hálft annað hundrað skipa. (AdvP-V-TOPIC)
then have.PST he.NOM half other hundred ships.GEN
'Then he had half of another hundred ships.' (1275, Morkin.268)

V1 sentences which have an immediately postfinite topic are also robustly attested, e.g. (4). This is the so-called 'narrative inversion' construction (Platzack, 1985; Sigurðsson, 1990).

[1]All examples come from the Icelandic Parsed Historical Corpus ('IcePaHC', Wallenberg et al., 2011) and are referenced in the form: Year, Text.UniqueID.

[2]Glossing throughout follows the Leipzig Glossing Rules: https://www.eva.mpg.de/lingua/resources/glossing-rules.php. In addition, we use EXPL to gloss an expletive (i.e. non-referential) *það*.

(4) a. Þórir **hann** þá eigi að stefna til gatnanna. (V-TOPIC)
dare.PRS he.NOM then NEG to go.INF to paths.DEF
'He then dares not make for the paths.' (1250, Sturlunga.445.2015)

b. Var **þetta smíði** hið virðulegasta. (V-TOPIC)
be.PST DEM.NOM building.NOM DEF magnificent.SUPL
'This building was the most magnificent.' (1400, Viglundur.94)

Moreover, many of these structures are still possible in the modern language (i.e. are attested in 1901-2008), e.g. (5).

(5) a. Sá **ég** þá á svipstundu villu míns vegar. (V-TOPIC)
see.PST I.NOM then on moment error my.GEN way.GEN
'I then saw in a moment the error of my way.' (1985, Margsaga.689)

b. Ekki mátti **saumavélin** til dæmis sigla... (NEG-V-TOPIC)
NEG could sewing-machine.DEF for example sail.INF
'The sewing machine could not for example sail' (1985, Margsaga.406)

c. Þá deyr **hann**. (AdvP-V-TOPIC)
then die.PRS he.NOM
'Then he dies.' (1920, Arin.1021)

Another striking observation from the historical data is that Icelandic has a small class of 'discourse adverbs' (DAs) which behave in an interesting way diachronically: *nú* 'now', *síðan* 'then', *svo* 'so', *þar* 'there', *þá* 'then'. These DAs can occur in the postfinite domain where they appear to separate TOPIC from FOCUS, e.g. (6).

(6) a. Þiggja þau **þar** ágærar gjafir. (V-TOPIC-DA-FOCUS)
receive.PRS they.NOM there excellent gifts
'They receive there excellent gifts.' (1350, Finnbogi.661.2086)

b. Konungurinn lá **þá** í Sólundum... (TOPIC-V-DA-FOCUS)
king.NOM.DEF lie.PST then in Sólundur
'The king was then at Sólundur.' (1260, Jomsvikingar.862)

Furthermore, DAs can also introduce the focused element in V1 sentences which lack a TOPIC (presentationals), e.g. (7).

(7) Voru **þar** tvö skip í búnaði. (V-DA-FOCUS)
be.PST there two.NOM ships.NOM in preparations
'There were two ships in the preparations.' (1250, Sturlunga.408.710)

The possibility that DAs serve an information-structural purpose has not been discussed for Icelandic. Parallel DAs have, however, been discussed for Early English, where it has been claimed that they function as discourse partitioners (van Kemenade & Los, 2006; van Kemenade, 2009).

In this paper, we show how this additional diachronically-informed data can be incorporated into a revised LFG analysis of Icelandic clause structure. Our analysis is informed by a series of corpus-based investigations using the Icelandic Parsed Historical Corpus ('IcePaHC', Wallenberg et al., 2011), which examine verb position, the positional distribution of topics and the positional distribution of DAs. We show that the c-structure configuration of Icelandic remains stable over time. What changes, however, is the way in which information structure is encoded syntactically, i.e. the association between information structure and c-structure, as captured via changing functional annotations on the c-structure.

2 Theoretical assumptions

We follow the standard LFG assumption that a functional category at c-structure is only motivated when functional information is associated with a fixed structural position (e.g. Kroeger, 1993; Börjars et al., 1999). Furthermore, unlike some analyses of Germanic V2 which posit extra layers of structure to account for functional information, we take the view that c-structure positions are only motivated via direct structural evidence (e.g. word order diagnostics, constituency tests); functional differences are sufficiently captured via functional annotations on the c-structure.

Our paper deals with information structure, where terminology is notoriously problematic. We take a feature-based approach and follow the four-way division of information structure by Butt & King (1996, 1997) shown in (8) (based on ideas from Vallduví 1992; Choi 1999; see also Mycock 2013; Butt et al. 2016).

(8)

	[+New]	[−New]
[+Prominent]	FOCUS	TOPIC
[−Prominent]	Completive information	BACKGROUND

In this view, FOCUS, TOPIC, and BACKGROUND are GDFs. Completive information is not especially salient, nor associated with a fixed structural position (see Butt & King, 1997), hence not a GDF. In Krifka (2007), FOCUS is defined as indicating the presence of alternatives relevant for the interpretation of a linguistic expression. As such, the features [+New] and [+Prominent] are correlates of this definition, see (8). We do not discuss contrastive focus in this paper, since contrastive foci occur relatively rarely in our historical corpus data.

TOPICS point to the entity about which relevant information should be stored in the Common Ground (Krifka, 2007). TOPICS thus signal what the expression is about (see also Butt et al., 2016). In this paper, we discuss continuing topics (e.g. Frascarelli & Hinterhölzl, 2007; cf. also center continuation in centering theory, Grosz et al., 1995). Hence, for us, TOPICS are [−New] and [+Prominent]. BACKGROUND material provides information as to how new information fits in with known information, i.e. the information necessary to provide a good understanding of the new (focused) information (Butt & King, 1997).

3 V1, V2 and I in Old Icelandic

Old Icelandic differs from other early Germanic languages in terms of verb position. Whereas Old English, Old High German and Old Saxon exhibit V1, V2, V3 and V-later structures (e.g. Kiparsky, 1995; Axel, 2007; Walkden, 2015), Old Icelandic is more restricted, having only V1 and V2. V3 or V-later structures do not occur in Old Icelandic (Faarlund, 1994, 64; Rögnvaldsson, 1995), as confirmed by a recent corpus study (Booth, 2018).

3.1 Data

V2 is robustly attested in Old Icelandic matrix clauses; the corpus study by Booth (2018) found that 82% of all matrix declaratives are V2. Old Icelandic V2 is illustrated by the examples in (9).

(9) a. Hann **átti** konu unga og fríða. (SUBJ-V-OBJ)
he.NOM own.PST woman.ACC young.ACC and beautiful.ACC
'He was married to a young and beautiful woman.' (1310, Grettir.312)

b. Hana **átti** Gamli Þórhallsson... (OBJ-V-SUBJ)
she.ACC own.PST Gamli.NOM Þórhallsson.NOM
'To her was married Gamli Þórhallsson...' (1310, Grettir.15)

c. Þar **átti** hann heima í Haugatungu. (ADJ-V-SUBJ)
there own.PST he.NOM home in Haugatungu
'He had home there at Haugatunga.' (1250, Sturlunga.389.30)

Strikingly, V1 is also exhibited in Old Icelandic matrix declaratives, i.e. beyond the typical contexts for V1 in modern Germanic (yes/no-interrogatives, imperatives). In the corpus study by Booth (2018), 18% of all matrix declaratives were V1 (Booth, 2018; see also Butt et al., 2014; Faarlund, 2004; Platzack, 1985; Sigurðsson, 1990; Walkden, 2014). Essentially, V1 declaratives fall into 3 types in Icelandic (Booth, 2018; Sigurðsson, 2018): (i) impersonal V1 (subjectless), e.g., (10-a); (ii) presentational V1 with a postfinite subject in focus, e.g. (10-b); (iii) narrative inversion V1, e.g. (10-c).

(10) a. **Tekur** nú að hausta. (impersonal)
begin.PRS now to become-autumn.INF
'It now starts to become autumn.' (1310, Grettir.48)

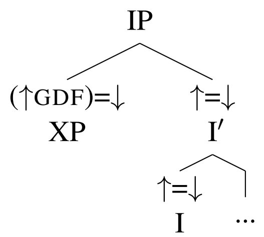
b. **Eru** nú hér með oss margir tígnir menn
be.PRS now here with we.ACC many.NOM noble.NOM men.NOM
og góðir drengir... (V-(...)-SUBJ$_{\text{FOCUS}}$)
and good.NOM boys.NOM
'There are now here with us many noble men and good boys...'
(1275, Morkin.401)

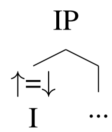
c. **Þórir** hann þá eigi að stefna til gatnanna. (V-SUBJ$_{\text{TOPIC}}$)
dare.PRS he.NOM then NEG to go.INF to paths.DEF
'He then dares not make for the paths.' (1250, Sturlunga.445.2015)

3.2 Analysis

Since Icelandic has only V1/V2 and no V3/V-later structures, we assume that I is already an obligatory functional category in Old Icelandic; a fixed structural position for finiteness. The proposal therefore is that all matrix declaratives in Old Icelandic are rooted in I, with one available specifier position (SpecIP). Our account for V1 and V2 – where they are rooted in the same c-structure – is in line with Kiparsky (1995). Kiparsky (1995) also argues that the functional head which hosts the finite verb is obligatory in Old Icelandic and optional in Old English, which accounts for the absence of V-later structures in the former and the presence in the latter. In Old English where I is optional, I is present in V1 and V2 sentences, but absent in V3 and V-later sentences.

In Old Icelandic V2 declaratives, SpecIP can be occupied by various categories e.g. subjects, objects, adjuncts, and is associated with a GDF. As such we propose (11) as the preliminary c-structure for V2 declaratives.

(11)


(12)


We assume that V1 declaratives in Old Icelandic are also IPs, where SpecIP is unoccupied, see (12) (see also Rögnvaldsson & Thráinsson, 1990 for modern Icelandic; Brandt et al., 1992 and Önnerfors, 1997 for German). This is contra other accounts for Germanic which assume that the finite verb is in C in V1 declaratives ('double verb movement' e.g. Sigurðsson, 1990 and Franco, 2008 for Icelandic; Mörnsjö, 2002 on Swedish). These accounts are motivated by the assumption that SpecIP is a unique subject position, within a framework where subjects are exclusively defined via structural position. In LFG, subjects are captured at f-structure and need not be associated with a fixed structural position at c-structure (Dalrymple, 2001). As we argue in Section 4, in Old Icelandic subject topics are not associated with a unique structural position. Our overall point is that in an LFG account we can adequately capture the various structural configurations in Icelandic matrix declaratives by assuming a c-structure rooted in I, without resorting to an additional CP-layer.

4 Topics in Old Icelandic

In Old Icelandic, topics may occur in the clause-initial position, see (13), as well as in the postfinite domain, see (14). That is, while topics can be placed in SpecIP, they may alternatively occur in the midfield. Postfinite topics occur, for example, when a non-topical element occupies SpecIP, see (14-a) with the DA *þá* 'then' in SpecIP, and (14-b), where an adjective is stylistically fronted to SpecIP.[3] Moreover, narrative inversion V1 clauses have a postfinite topic, see (10-c).

(13) TOPIC-V

a. **Hann** átti konu unga og fríða.
he.NOM own.PST woman.ACC young.ACC and beautiful.ACC
'He was married to a young and beautiful woman.'
(1310, Grettir.312)

b. **Hana** átti Gamli Þórhallsson...
she.ACC own.PST Gamli.NOM Þórhallsson.NOM
'To her was married Gamli Þórhallsson...' (1310, Grettir.15)

c. **Öxin** kom á herðarblaðið.
OX.NOM.DEF come.PST on shoulder-blade.DEF
'The ox came up onto his shoulder blade.' (1310, Grettir.1120)

(14) XP-V-TOPIC

a. Þá hafði **hann** hálft annað hundrað skipa.
then have.PST he.NOM half other hundred ships.GEN
'Then he had half of another hundred ships.' (1275, Morkin.268)

b. En fullt var **skipið**.
but full.NOM be.PST ship.NOM.DEF
'But the ship was full.' (1210, Jartein.779)

In order to test whether topics prefer a particular position in Old Icelandic, we conducted a corpus study investigating their positional distribution in matrix declaratives in texts from IcePaHC which date from 1150 to 1350, i.e., cover the Old Icelandic period. The results of this study are reported in the following.

4.1 Corpus study

IcePaHC does not annotate for information structure. Thus, as an approximation of topics, we took any referential NP argument which is pronominal or has overt definite marking, since these properties can be extracted from the annotation.[4] More-

[3] We assume that these examples are cases of 'Stylistic Fronting' (fronting of categories which cannot usually be fronted in a Germanic V2 language), even though they flaunt the 'subject gap condition' established for modern Icelandic (Maling, 1990).

[4] We are well aware of the limitations to this approach: (i) not all pronominal and definite NP arguments will be topics and (ii) definiteness marking was not yet obligatory for semantically definite NPs in Old Icelandic, so we will not have captured all semantically definite NPs. However, keeping this in mind, we are convinced that the data provided here still gives valuable clues about the po-

over, we decided to focus on topical subjects (and not e.g. topical objects) in order to make the corpus investigation of manageable scope.

The results of our investigation into the positional distribution of topics in Old Icelandic are given in Table 1. We calculated the occurrence frequencies of the following constructions: (i) V2 clauses which have a topic in SpecIP (TOPIC-V), (ii) narrative inversion V1 clauses with a postfinite topic (V-TOPIC), (iii) V2 clauses with a DA in SpecIP and a postfinite topic (DA-V-TOPIC) and (iv) V2 clauses with a stylistically fronted element in SpecIP and a postfinite topic (SF-V-TOPIC).[5]

TOPIC-V		V-TOPIC		DA-V-TOPIC		SF-V-TOPIC	
n	%	*n*	%	*n*	%	*n*	%
1574	58.8%	679	25.4%	381	14.2%	43	1.6%

Table 1: Positional distribution of topical subjects in Old Icelandic (1150-1350).

The data in Table 1 shows that topics occur preferably in the clause-initial position, i.e., SpecIP in Old Icelandic (58.8%). However, topics also regularly appear postverbally, in particular in the context of V1, and to a lesser extent with the clause-initial DAs. SF with a postfinite topic is comparably rare in Old Icelandic, but then SF is a rare phenomenon overall.[6]

Altogether, this indicates that the functionally annotated c-structures postulated for modern Icelandic, i.e., (1) and (2) (Sells, 2005; Booth et al., 2017), cannot hold for Old Icelandic. While it has been argued that SpecIP is the topic position in the modern language, the data presented here shows that topics additionally occur regularly in the midfield in Old Icelandic. These results now inform the functional annotations which we add to the c-structure skeletons given in Section 3.2 in (11) and (12).

4.2 Analysis

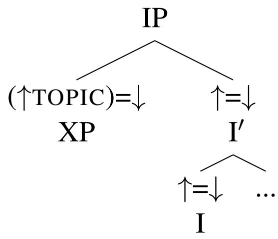
In essence, the corpus study on the positional distribution of topics in Old Icelandic has shown that there are two possible topic positions. The preferred position is the clause-initial prefinite position (SpecIP). Accounting for this, we posit the c-structure tree in (15) for Old Icelandic. Another possibility is for topics to occur in the immediately postfinite position, i.e., in the midfield under I′. This applies to V1 clauses (narrative inversion), where SpecIP remains unoccupied, for which we give the structure in (16).

sitional distribution of topics in Old Icelandic and see this as a promising starting point for further investigations.

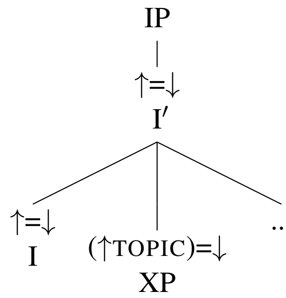
[5]As an approximation of Stylistic Fronting we count matrix declaratives where a nonfinite verb, verbal particle, negation, an adjectival or nominal predicate occurs in SpecIP (e.g. Maling, 1990).

[6]Out of a total 19,771 matrix clauses in IcePaHC for the period 1150-1350, 160 have an SF element in SpecIP by our criteria (0.8%).

(15)

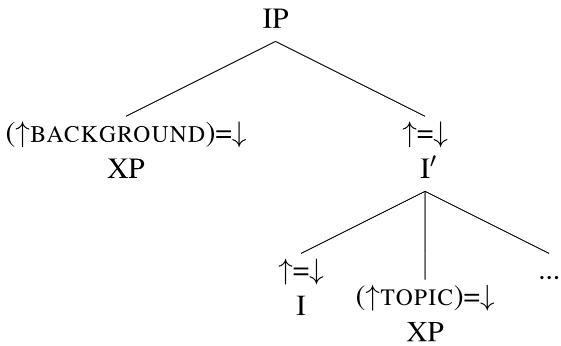


(16)



Furthermore, for clauses with a postfinite topic, e.g. (14-a), where a DA sits in SpecIP, we assume the structure in (17).

(17)



In structures like (17) with a postfinite TOPIC ([–New, +Prominent]), we suggest that SpecIP is a 'discourse-linking' position, i.e. can host BACKGROUND information which is [–New, –Prominent]. This accounts for the fact that DAs (ADJs) commonly occur clause-initially, even in clauses with topics.

Furthermore, this proposal fits with our analysis of narrative inversion V1, see the c-structure in (16), since this type of V1 is limited to contexts where the same scene is maintained and there is no need for a scene-setter or discourse-linker in SpecIP. This is illustrated by the continuous narrative in (18), where a V2 clause indicates a scene change, while a V1 clause occurs when the same scene is continuously maintained.

(18) Continuous narrative:

a. Gissur **kom** í Reykjaholt um nóttina eftir Máritíusmessu.
Gissur come.PST to Reykjaholt in night.DEF after mass
'Gissur came to Reykjaholt in the night after the mass.'
(scene change > V2)

b. **Brutu** þeir upp skemmuna er Snorri svaf í.
break.PST they.NOM up storehouse.DEF REL Snorri sleep.PST in
'They broke open the storehouse where Snorri was sleeping.'
(same scene > V1)

c. En hann **hljóp** upp og úr skemmunni og í hin litlu
but he.NOM leap.PST up and out storehouse.DEF and in DEF little
húsin er voru við skemmuna.
houses.DEF REL be.PST by storehouse.DEF
'But he leaped up and out of the storehouse and into the little houses which were next to the storehouse.' (scene-change > V2)

d. **Fann** hann þar Arnbjörn prest og talaði við hann.
find.PST he.NOM there Arnbjörn priest and speak.PST with he.ACC
'He found there Arnbjörn the priest and spoke with him.'
(same scene > V1)

e. **Réðu** þeir það að Snorri gekk í kjallarann er
plan.PST they.NOM DEM COMP Snorri go.PST in cellar.DEF REL
var undir loftinu þar í húsunum.
be.PST under ceiling.DEF there in house.DEF
'They plotted that Snorri would go into the cellar which was under the ceiling there in the house.' (same scene > V1)

f. Þeir Gissur **fóru** að leita Snorra um húsin.
they.NOM Gissur begin.PST to lead.INF Snorri around house.DEF
'They and Gissur began to lead Snorri around the house.'
(scene change > V2)
(1250, Sturlunga.439.1766 – 1250, Sturlunga.439.1772)

Moreover, the account is in line with the fact that 'out-of-the-blue' presentationals are typically V1, e.g. (10-b). V1 presentationals in Icelandic are 'all new' sentences. Thus, there is no motivation for overt discourse-linking, i.e., BACKGROUND material in the clause-initial position, and SpecIP remains unoccupied.

Additionally, the blueprint in (17) could also work with Stylistic Fronting if we follow the proposal by Egerland (2013) that SF is a backgrounding device. However, more in-depth research on the nature of SF in historical Icelandic is necessary to be able to draw more definite conclusions with respect to its information-structural impact.

If we interpret the c-structures in (15) and (17), where the SpecIP position is filled, in terms of the feature space given for information structure in (8), we arrive at the structure in (19). Given the possibility for both TOPIC and BACKGROUND to occur in SpecIP, we characterise this position as [−New].[7]

[7] Although not made explicit in the present analysis, we generally assume that information-structural content is projected to a separate i(nformation)-structure (following e.g. Butt et al., 2016).

(19)

```
                    IP
           ________/  \________
  (↑GDF)=↓                     ↑=↓
  (↑GDF TYPE NEW = −)          I′
                              /  \
                            ↑=↓   ...
                             I
```

In sum, investigating and understanding the behaviour of topics has given us insights into both SpecIP and the midfield, as well as into discourse management strategies. Next, we investigate the positional distribution of discourse adverbs in Old Icelandic further, in order to yet broaden our understanding of the interrelation between information structure and clause structure.

5 Discourse adverbs in Old Icelandic

DAs can occur in various positions in Old Icelandic. Firstly, they can occur in the clause-initial prefinite position (SpecIP), e.g. (20). However, DAs also commonly occur in the midfield, e.g. (21).

(20) **Þá** hafði hann hálft annað hundrað skipa. (DA-V)
then have.PST he.NOM half other hundred ships.GEN
'Then he had half of another hundred ships.' (1275, Morkin.268)

(21) a. Konungurinn lá **þá** í Sólundum... (TOPIC-V-DA)
king.DEF lie.PST then in Sólundur
'The king was then at Sólundur.' (1260, Jomsvikingar.862)

b. Þiggja þau **þar** ágærar gjafir. (V-TOPIC-DA)
receive.PRS they.NOM there excellent.ACC gifts.ACC
'They receive there excellent gifts.' (1350, Finnbogi.661.2086)

c. Voru **þar** tvö skip í búnaði. (V-DA-FOCUS)
be.PST there two.NOM ships.NOM in preparations
'There were two ships in the preparations.' (1250, Sturlunga.408.710)

Parallel DAs in Early English have been claimed to serve an information-structural role, separating topic from focus (van Kemenade & Los, 2006; van Kemenade, 2009). The behaviour of DAs in historical Icelandic has scarcely been investigated to date. In this section we examine their positional distribution in Old Icelandic via a corpus-based study.

5.1 Corpus study

We examine the relative frequencies at which DAs occur across the 4 different configurations in (20)-(21) for the Old Icelandic texts in IcePaHC (1150-1350). The findings are shown in Table 2.

DA-V		TOPIC-V-DA		V-TOPIC-DA		V-DA-FOCUS	
n	%	*n*	%	*n*	%	*n*	%
1001	72.0%	196	14.1%	185	13.3%	9	0.6%

Table 2: Positional distribution of discourse adverbs in Old Icelandic (1150-1350).

The key observation is that while DAs are predominantly prefinite, i.e. occur in SpecIP (DA-V, 72.0%), they also occur in the postfinite domain, i.e. midfield (columns 2-4). In the postfinite domain, DAs occur at a comparable frequency in V2 topic-initial clauses (14.1%) as in V1 narrative inversion (13.3%). DAs do occur in V-DA-FOCUS structures (presentationals), but presentationals overall are rare in the corpus data (see Booth, 2018).

5.2 Analysis

We propose that DAs (which are [–New] and [–Prominent] in information-structural terms) in SpecIP are discourse-linkers; they qualify as BACKGROUND, see (8) above. This is in line with previous work on early Germanic: see van Kemenade & Los (2006) on 'discourse operators' in Early English; Hinterhölzl & Petrova (2010) and Petrova & Rinke (2014) on 'discourse linkers'/'discourse-linking elements' in Old High German (and Old French); Los (2009) and Komen et al. (2014) on the discourse-linking function of the clause-initial position in Old and Middle English.

We claim that the DAs which occur in the midfield can also serve as an information-structural boundary separating TOPIC and FOCUS, see (21-a)-(21-b) (cf. van Kemenade & Los, 2006 on Early English). In line with this role as a discourse partitioner, a midfield DA in V1 sentences which lack a topic (presentationals) closes off the (empty) topic domain and introduces the focus, see (21-c). This is contra the previous proposal for Icelandic by Booth et al. (2017), in which the finite verb (in I) is assumed to be an information-structural boundary separating topic (prefinite) and comment (postfinite) (see also Hinterhölzl & Petrova, 2010 on Early West Germanic). Having I as an information-structural boundary closing off the topic domain clearly does not work for Old Icelandic, where topics occur relatively frequently in the postfinite domain (see Section 4).

Thus we propose that DAs can serve two different information-structural roles in Old Icelandic, which correlate with structural position:

(22) a. discourse linker: DA-V-TOPIC
b. discourse partitioner:
(i) TOPIC-V-DA-FOCUS
(ii) V-TOPIC-DA-FOCUS
(iii) V-DA-FOCUS

We wish to point out that we are not claiming here that DAs are fully grammaticalised elements. We assume that they retain their temporal-spatial semantics, but have taken on an additional layered information-structural function. We shed more light on the historical development of DAs in the next section.

6 Continuity and change

In the previous sections, we have provided an account for Old Icelandic clause structure on the basis of data from IcePaHC. Now, we turn to how this account can be reconciled with the previous LFG accounts of modern Icelandic clause structure (Sells, 2005; Booth et al., 2017) in terms of continuity and change across the Icelandic diachrony. For this purpose, we investigate the diachronic interaction between information structure and word order over the nine centuries of Icelandic spanned by IcePaHC (1150-2008). To assess the historical developments, we divide the corpus data into periods which have been derived via a data-driven method for periodisation using hierarchical clustering, i.e., DiaHClust (see Schätzle & Booth 2019 for details). Via this method, the IcePaHC texts are grouped into time stages based on their similarity with respect to known changing syntactic features. This results in the following time stages: 1150-1210, 1250-1450, 1475-1630, 1650-1882, 1883-2008.[8] Moreover, this method carves out the genre bias inherent in texts around the 16th century (bible translations stemming from the Reformation) which was previously identified and is known to affect the syntactic characteristics (e.g. Butt et al., 2014; Booth et al., 2017), clustering these texts into the period 1475-1630.

We moreover compute χ^2-tests to calculate whether the observed distributions in each period differ significantly from what could be expected given the total number of data points in each period and the overall distributions of the individual constructions across all periods. In this way, we compare the actual distributions with the distributions that would occur if the data were equally distributed across periods (given the number of data points in each period). Significant differences between the observed and the expected distributions are thus taken to be indicative of change (with $p < 0.05$ '*', $p < 0.01$ '**', $p < 0.001$ '***').

6.1 Continuity

It has been shown in earlier work that V2 is robustly and continuously attested in matrix clauses throughout the Icelandic diachrony (see Butt et al. 2014; Booth et al. 2017). V1 is still an option in the modern stage, although V1 decreases significantly in frequency over time. The results obtained by Booth et al. (2017) are given in Table 3, where the 'non V1' column in essence depicts the occurrence frequencies of V2 in IcePaHC.[9]

[8] The stages are discontinuous because they begin and end with the year date of the first, or respectively last, text belonging to the individual clusters.

[9] V3 order is possible at least in modern Icelandic with certain adverbs (Angantýsson, 2007, 241; Thráinsson, 2007, 22, 53) but it is a fringe phenomenon and does not show up in the earlier periods in IcePaHC, so we do not discuss it further here.

Period	**V1**	**non V1**	**Total**	**% V1**	χ^2
1150-1349	2829	10888	13718	20.6%	***
1350-1549	3656	14693	18349	19.9%	***
1550-1749	1654	9556	11210	14.8%	***
1750-1899	2072	9185	11257	18.4%	***
1900-2008	292	10569	10861	2.7%	***

Table 3: Distribution of V1 matrix declaratives in IcePaHC (Booth et al., 2017).

Taking this into account, we assume that the functional category I remains obligatory in Icelandic, consistently hosting the finite verb. Moreover, SpecIP is still optional in the modern language, as V1 clauses remain a part of the language. Yet, SpecIP is increasingly occupied over time and the frequency of V1 decreases.

6.2 Change

The IcePaHC data presented in this section show that the association between i-structure and c-structure changes over time in Icelandic. For one, topics increasingly target SpecIP (see also Booth et al., 2017 for quantitative evidence). SpecIP in turn is becoming more firmly associated with topics. This then allows the finite verb to serve as a boundary between the TOPIC (SpecIP) and the midfield. For this reason, DAs as discourse partitioners in the midfield are no longer motivated and we observe a striking decrease in midfield DAs. The respective corpus findings are detailed in the following.

6.2.1 Topics and SpecIP

Table 4 shows the positional distribution of topical subjects in IcePaHC (1150-2008). Again, we take any referential NP argument which is pronominal or has overt definite marking as an approximation for topics.

Period	TOPIC-V		V-TOPIC		DA-V-TOPIC		SF-V-TOPIC		χ^2
	n	%	*n*	%	*n*	%	*n*	%	
1150-1210	266	**49.4%**	130	24.2%	129	24.0%	13	2.4%	***
1250-1450	2014	**57.7%**	1031	29.5%	400	11.5%	47	1.3%	***
1475-1630	748	**71.5%**	69	6.6%	208	19.9%	21	2.0%	***
1650-1882	1795	**59.0%**	876	28.8%	316	10.4%	56	1.8%	***
1883-2008	2593	**88.3%**	76	2.6%	231	7.9%	37	1.3%	***

Table 4: Positional distribution of topical subjects in IcePaHC (1150-2008).

Overall, topics occur most frequently in the clause-initial position, i.e. SpecIP, see Table 4 (TOPIC-V). Furthermore, this preference increases diachronically, with up to 88.3% of the topics occurring in SpecIP in the period 1883-2008, while the constructions with a midfield topic decrease. Also, the increase in topics in SpecIP goes hand in hand with a decrease in DAs in that position (DA-V-TOPIC).

We explain this in terms of the changing information-structural associations of SpecIP. We have shown in Section 4 that SpecIP was associated with the information-structural feature [−New] in Old Icelandic, cf. the tree in (19). In modern Icelandic, referring to the period post-1883 in our data, SpecIP is still characterised as [−New] but increasingly hosts topics which, in addition to being [−New], are [+Prominent]. Thus, SpecIP becomes increasingly associated with prominence over time, resulting in the structure given in (23). Hence, the tendency for BACKGROUND material, e.g., DAs, which is [−Prominent], to occur in SpecIP recedes. Moreover, a further consequence of the increasing association of topics with SpecIP is the increased dominance of V2 over V1, leading to the gradual loss of V1 declaratives in the language.

(23) IP

(↑GDF)=↓ (↑GDF TYPE NEW = −) (↑GDF TYPE PROMINENT = +) | ↑=↓ I′

I′: ↑=↓ I | ...

Now that the topics in SpecIP are clearly demarcated from the rest of the clause via the finite verb in I, I can function as information-structural boundary separating topic and comment (i.e. completive information, see (8)) in the modern language. This is in line with Booth et al. (2017). Since we have shown earlier that DAs were functioning as discourse partitioners in Old Icelandic, we examine the diachrony of DAs more closely in the next section, investigating the trade-off between DAs and SpecIP in terms of the syntactic encoding of information structure.

6.2.2 Discourse adverbs

Table 5 displays the positional distribution of DAs across time (IcePaHC, 1150-2008). We have shown in Section 6.2.1 that DAs are a receding option for SpecIP in comparison to topics. However, when looking at the positional distribution of DAs alone, we found that DAs are in fact increasingly confined to SpecIP (DA-V), while the other placement options (midfield DA), decrease over time.

We account for the increasing confinement of DAs to SpecIP in terms of the increasing loss of their function as discourse partitioners in the midfield. Since topics are now more firmly associated with SpecIP, DAs as a discourse partitioner separating TOPIC from FOCUS are no longer motivated. This is moreover supported by the fact that the constructions which have a topic in SpecIP together with a midfield DA (TOPIC-V-DA), i.e., constructions where the verb already functions as a discourse partitioner and the DA is in principle redundant with respect to this

Period	DA-V		TOPIC-V-DA		V-TOPIC-DA		V-DA-FOCUS		χ^2
	n	%	*n*	%	*n*	%	*n*	%	
1150-1210	209	**77.7%**	23	8.6%	36	13.4%	1	0.4%	*
1210-1450	1191	**68.8%**	162	9.4%	359	20.8%	18	1.0%	***
1475-1630	495	**95.0%**	0	0.0%	22	4.2%	4	0.8%	***
1650-1882	788	**76.8%**	0	0.0%	232	22.6%	6	0.6%	***
1883-2008	368	**96.8%**	0	0.0%	12	3.2%	0	0.0%	***

Table 5: Positional distribution of discourse adverbs in IcePaHC (1150-2008).

function, are virtually lost already after the Old Icelandic period (post-1350).

Overall, we suggest that SpecIP is becoming a topic position. Moreover, since DAs are drastically reduced in the postfinite domain, they no longer function as a discourse partitioner, and the finite verb in I is taking over as an information-structural boundary, delimiting the topic domain, i.e., SpecIP.

7 Conclusion

In this paper we have shown that the same basic c-structure with I as a functional category persists throughout the Icelandic diachrony, while the association between c-structure and information structure changes. Topics become more firmly associated with SpecIP, allowing I to serve as a boundary between topic and comment. Connected with this change, discourse adverbs in the midfield no longer function as discourse partitioners, but rather mainly occur in SpecIP where they function as discourse-linkers.

Finally, this series of changes can be related to other syntactic developments previously shown for Icelandic. One of these developments is the increasing occurrence of the expletive *það* in SpecIP (see Booth et al. 2017, Booth 2018). Along with SpecIP becoming an established topic position, expletive *það* increases in frequency as a filler for this position and as a signaller of an 'all new' clause (Booth, 2018), e.g. (24).

(24) **Það** var töluverður snjór yfir öllu. (EXPL-V-FOCUS)
EXPL be.PST considerable.NOM snow.NOM over everything
'There was a considerable amount of snow over everything.'
(2008, Ofsi.772)

With the SpecIP expletive now an information-structural signal of a topicless sentence, midfield DAs in V1 presentationals such as in (21-c) are no longer motivated to close off the topic domain. This is further supported by the fact that presentational constructions which have both the expletive in SpecIP and a midfield DA are not attested in IcePaHC.

Another development which can be related to the increasing association of topics with SpecIP is that subjects increasingly occur in SpecIP. Since topics are often subjects, subjects overall increasingly target SpecIP and SpecIP is on its way to becoming a subject licensing position, as previously claimed (Booth et al., 2017; Schätzle, 2018).

References

Angantýsson, Ásgrímur. 2007. Verb-third in embedded clauses in Icelandic. *Studia Linguistica* 61(3). 237–260.

Axel, K. 2007. *Studies on Old High German Syntax: Left Sentence Periphery, Verb Placement and Verb-Second*. Amsterdam: John Benjamins.

Booth, Hannah. 2018. *Clause Structure and Expletives: Syntactic Change in Icelandic*. Ph.D thesis, University of Manchester.

Booth, Hannah, Christin Schätzle, Kersti Börjars & Miriam Butt. 2017. Dative subjects and the rise of positional licensing in Icelandic. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'17 Conference, University of Konstanz*, 104–124. Stanford, CA: CSLI Publications.

Börjars, Kersti, Elisabet Engdahl & Maia Andréasson. 2003. Subject and object positions in Swedish. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'03 Conference, University at Albany, State University of New York*, 43–58. Stanford, CA: CSLI Publications.

Börjars, Kersti, John Payne & Erika Chisarik. 1999. On the justification for functional categories in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'99 Conference, University of Manchester*, Stanford, CA: CSLI Publications.

Brandt, Margareta, Marga Reis, Inger Rosengren & Ilse Zimmermann. 1992. Satztyp, Satzmodus und Illokution. In Inger Rosengren (ed.), *Satz und Illokution*, vol. 1, 1–90. Tübingen: Niemeyer.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax, 2nd Edition*. Oxford: Blackwell.

Butt, Miriam, Tina Bögel, Kristina Kotcheva, Christin Schätzle, Christian Rohrdantz, Dominik Sacha, Nicole Dehé & Daniel A. Keim. 2014. V1 in Icelandic: A multifactorical visualization of historical data. In *Proceedings of the LREC 2014 Workshop on Visualization as added value in the development, use and evaluation of LRs (VisLR)*, 33–40. Reykjavík, Iceland.

Butt, Miriam, Farhat Jabeen & Tina Bögel. 2016. Verb Cluster Internal Wh-Phrases in Urdu: Prosody, Syntax and Semantics/Pragmatics. *Linguistic Analysis* 40(3-4). 445–487.

Butt, Miriam & Tracy Holloway King. 1996. Structural topic and focus without movement. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the First Annual LFG Conference*, Stanford: CSLI Publications.

Butt, Miriam & Tracy Holloway King. 1997. Null elements in discourse structure. In Karumuri V. Subbarao (ed.), *Papers from the NULLS seminar*, Delhi: Motilal Banarsidass.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI Publications.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.

Egerland, Verner. 2013. Fronting, background, focus: a comparative study of Sardinian and Icelandic. *Lingua* 136. 63–76.

Faarlund, Jan Terje. 1994. Old and Middle Scandinavian. In Johan van der Auwera & Ekkehard König (eds.), *The Germanic Languages*, 38–71. London: Routledge.

Faarlund, Jan Terje. 2004. *The Syntax of Old Norse: with a survey of the inflectional morphology and a complete bibliography*. Oxford: Oxford University Press.

Franco, Irene. 2008. V1, V2 and criterial movement in Icelandic. *Studies in Linguistics* 2. 141–164.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Susanne Winkler & Kerstin Schwabe (eds.), *On Information Structure, Meaning and Form*, 87–116. Amsterdam: John Benjamins.

Grosz, Barbara J., Scott Weinstein & Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21. 203–225.

Hinterhölzl, Roland & Svetlana Petrova. 2010. From V1 to V2 in West Germanic. *Lingua* 120(2). 315–328.

van Kemenade, Ans. 2009. Discourse relations and word order change. In Roland Hinterhölzl & Svetlana Petrova (eds.), *Information Structure and Language Change: New Approaches to Word Order Variation in Germanic*, 91–117. Berlin: de Gruyter.

van Kemenade, Ans & Bettelou Los. 2006. Discourse adverbs and clausal syntax in Old and Middle English. In Ans van Kemanade & Bettelou Los (eds.), *The Handbook of the History of English*, 224–248. Oxford: Blackwell.

Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In Ian Roberts & Adrian Battye (eds.), *Clause Structure and Language Change*, 140–167. Oxford: Oxford University Press.

Komen, Erwin R., Rosanne Hebing, Ans van Kemenade & Bettelou Los. 2014. Quantifying information structure change in English. In *Information Structure and Syntactic Change in Germanic and Romance Languages*, 81–110. Amsterdam/Philadelphia: John Benjamins.

Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry & Manfred Krifka (eds.), *Interdisciplinary Studies on Information Structure*, vol. 6, 13–56. Potsdam: Universitätsverlag.

Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

Los, Bettelou. 2009. The consequences of the loss of verb-second in English: Information structure and syntax in interaction. *English Language and Linguistics* 13(1). 97–125.

Maling, Joan. 1988. Variations on a theme: existential sentences in Swedish and Icelandic. *McGill Working Papers in Linguistics: Special Issue on Comparative Germanic Syntax* 6(1). 168–191.

Maling, Joan. 1990. Inversion in embedded clauses in Modern Icelandic. In Joan Maling & Annie Zaenen (eds.), *Syntax and Semantics: Modern Icelandic Syntax*, 71–91. San Diego, CA: Academic Press.

Mycock, Louise. 2013. Discourse functions of question words. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'13 conference*, 419–439. Stanford, CA: CSLI Publications.

Mörnsjö, Maria. 2002. *V1 Declaratives in Spoken Swedish: Syntax, Information Structure, and Prosodic Pattern*. Doctoral dissertation, Lund University.

Önnerfors, Olaf. 1997. On narrative declarative V1 sentences in German. In Toril Swan & Olaf Jansen Westvik (eds.), *Modality in Germanic Languages: Historical and Comparative Perspectives*, 293–319. Berlin: de Gruyter.

Petrova, Svetlana & Esther Rinke. 2014. Formal properties of event-reporting sentences in Old High German and Old French. In *Information Structure and Syntactic Change in Germanic and Romance Languages*, 271–294. Amsterdam/Philadelphia: John Benjamins.

Platzack, Christer. 1985. Narrative inversion in Old Icelandic. *Íslenskt mál* 7. 127–144.

Rögnvaldsson, Eiríkur. 1995. Old Icelandic: a non-configurational language? *NOWELE* 26. 3–29.

Rögnvaldsson, Eiríkur & Höskuldur Thráinsson. 1990. On Icelandic word order once more. In Joan Maling & Annie Zaenen (eds.), *Syntax and Semantics: Modern Icelandic Syntax*, 3–40. San Diego, CA: Academic Press.

Schätzle, Christin. 2018. *Dative subjects: Historical change visualized*: University of Konstanz dissertation.

Schätzle, Christin & Hannah Booth. 2019. DiaHClust: an iterative hierarchical clustering approach for identifying stages in language change. In *Proceedings of the 1st international workshop on computational approaches to historical language change*, 126–135. Florence, Italy: Association for Computational Linguistics. https://www.aclweb.org/anthology/W19-4716.

Sells, Peter. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford, CA: CSLI publications.

Sells, Peter. 2005. The peripherality of the Icelandic expletive. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'05 Conference, University of Bergen*, 408–428. Stanford, CA: CSLI Publications.

Sigurðsson, Halldór Ármann. 1990. V1 declaratives and verb raising in Icelandic. In Joan Maling & Annie Zaenen (eds.), *Syntax and Semantics: Modern Icelandic Syntax*, 41–69. San Diego, CA: Academic Press.

Sigurðsson, Halldór Ármann. 2007. Argument features, clausal structure and the computation. In Tammoy Bhattachayra, Eric Reuland & Giorgos Spathas (eds.), *Argument Structure*, 121–158. Amsterdam: John Benjamins.

Sigurðsson, Halldór Ármann. 2018. Icelandic declarative V1: a brief overview. *Working Papers in Scandinavian Syntax* 49–55.

Thráinsson, Höskuldur. 1979. *On Complementation in Icelandic*. New York: Garland.

Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.

Vallduví, Enric. 1992. *The Informational Component*. New York: Garland Press.

Walkden, George. 2014. *Syntactic Reconstruction and Proto-Germanic*. Oxford: Oxford University Press.

Walkden, George. 2015. Verb-third in early West Germanic: a comparative perspective. In Theresa Biberauer & George Walkden (eds.), *Syntax over Time: Lexical, Morphological and Information-Structural Interactions*, 236–248. Oxford: Oxford University Press.

Wallenberg, Joel C., Anton Karl Ingason, Einar Freyr Sigurðsson & Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC), version 0.9. http://linguist.is/icelandic_treebank.

A Non-Canonical Diachronic Formation of Raising Predicates

Maris Camilleri
University of Essex

Louisa Sadler
University of Essex

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 90–110

http://csli-publications.stanford.edu/LFG/2019

Keywords: Arabic, raising, dialects, historical change, verbs of perception

Camilleri, Maris, & Sadler, Louisa. 2019. A Non-Canonical Diachronic Formation of Raising Predicates. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 90–110. Stanford, CA: CSLI Publications.

Abstract

While copy raising structures have been discussed for Arabic, it is sometimes claimed in the literature that SUBJ-to-SUBJ raising in the context of lexical predicates such as perceptual report verbs does not occur. With data from an array of vernaculars we argue that such structures do exist, and we demonstrate how they do not necessarily involve what we would take to be typical lexical verbal predicates. SSR constructions are more likely to be expressed through the use of non-canonical predicates. We discuss the grammaticalisation path of the N *šakl* 'form, shape', and the P 'like, as' which we hypothesise to have led to their development as verbal perceptual report predicates, and ones which appear in SSR constructions. We argue that the presence of a Prominent Internal Possessor is an enabling factor in the diachronic development of *šakl*. As for the P 'like', in addition to its complementising role in marking copy-raising predicates in some varieties of Arabic, we suggest that it has also emerged as a perceptual report predicate in its own right in at least one dialect.

1 Introduction

This paper looks at lexical elements which we argue have grammaticalised into raising predicates of a particular sort in the Arabic dialects. In doing so we aim to contribute to (i) grammaticalisation in identifying non-canonical sources of raising predicates, and (ii) to the study of the synchronic syntax of verb complementation in Arabic, where much previous work, especially on control and raising, has been concerned with Modern Standard Arabic (MSA) and has focussed on rather typical (crosslinguistically canonical) instances/sources of raising predicates. We here particularly focus on the vernacular varieties, rather than MSA, and focus our attention on the grammaticalisation of **lexical** raising predicates, rather than the formation of raising constructions involving TAM-type auxiliaries. Here we highlight two routes/sources that have led to the formation of lexical SUBJ-to-SUBJ raising structures from unusual sources, and we use LFG to encode our hypotheses about the diachronic changes which must have taken place.

2 Raising

2.1 Background

In its core instances, a raising construction involves a predicate that occurs with a non-thematic (term) syntactic argument (SUBJ or OBJ) which is a thematic argument (a SUBJ) of an embedded predication (an XCOMP). The relationship between the 'higher' (non-thematic) syntactic argument and the 'lower' (thematic,

†This work was partially funded by a Leverhulme Major Research Fellowship MRF-2016-048. Support from this source is gratefully acknowledged. Thanks also to the reviewers for very useful comments and feedback during the reviewing process.

except in cases of chained raising predicates), is expressed in a functional control equation such as ($\uparrow$ SUBJ) = ($\uparrow$ XCOMP SUBJ), which in turn accounts for the structure in (1a), as in (1a). Typically, a raising predicate will also allow non-raised truth-conditionally equivalent constructions in (1c) in which a dummy or expletive element occurs as the non-thematic syntactic argument. In these instances, the equation in (1b) does not figure in (1d). Predicates with non-thematic subjects do not always permit raising alternatives (English *probable* does not, for example).

(1) a. Chris seemed to enjoy the marathon. SSR

b. *seem* ($\uparrow$ PRED) = 'SEEM< XCOMP> SUBJ'
($\uparrow$ SUBJ) = ($\uparrow$ XCOMP SUBJ)

c. It seems that Chris enjoyed the marathon. *it*-expletive structure

d. *seem* ($\uparrow$ PRED) = 'SEEM< COMP> SUBJ'

However, apart from structure-sharing, as in (1b), raising structures may alternatively involve the mechanism of anaphoric binding (Asudeh and Toivonen, 2012) in copy raising (CR) constructions such as (2a), in which a resumptive pronoun occurs in the 'lower', thematic argument position (see Asudeh (2012) and Asudeh and Toivonen (2012) for a semantic analysis in which the additional resource contributed by the pronominal is managed away in semantic composition). Copy-raising predicates such as English *seem* or *appear* also have expletive (non copy-raised) counterparts, as in (2b-c).

(2) a. Chris$_i$ seemed as if/like/as though he$_i$ enjoyed the marathon.
CR: Asudeh and Toivonen (2012, 120)

b. It seems like Harry fell. Asudeh (2012, 328)

c. It appears as if Alfred hurt Harry. Asudeh (2012, 328)

Beyond the presence of a pronoun in the embedded clause, which gives the name to this construction, two key characteristics of copy raising are (i) the occurrence of a *like, as though, as if* complement which mediates the relationship between the raising predicate and the embedded predication, where *like*, and *as though* are analysed as prepositional in Maling (1983); and (ii) an entailment that the SUBJ in the matrix is understood as an (individual) perceptual source (PSOURCE), i.e. where it is something about the very nature of the SUBJ, rather than anything in the eventuality that is what helps us infer the proposition in the embedded clause. Note that PSOURCE is merely a semantic role, and not a thematic role, and indeed Asudeh (2012) takes examples such as (2b-c) as empirical evidence that the raising predicate does not take a thematic subject.

2.2 Raising and Arabic

Discussions of raising in Arabic mainly focus on Modern Standard Arabic (MSA), although they often make somewhat sweeping statements about Arabic varieties in general. There would appear to be no general consensus: authors such as Mohammad (2000) and Soltan (2007) argue that raising does not occur, while others such as Salih (1985) claim that it does. The arguments against the availability of raising come mainly from the fact that verbs such as *yabdū* and *yaḏhar* 'seem', which are typically raising predicates crosslinguistically, are said to always take an expletive subject expressed as default 3SGM morphology on the verb, along with a sentential complement introduced by the complementiser *ʔanna*+ACC pronoun.

Evidence that (3a) corresponds to an *it*-expletive construction, where the matrix involves a default 3SGM form of the verb rather than agreement with the SGM 'teacher' comes from the fact that changing the embedded SUBJ to SGF does not permit 3SGF agreement to appear on the matrix *bada* 'seem'.

(3) a. **ya-bdu** ʔanna l-muʕallim-a
3SM-seem.IPFV.SG.INDIC that DEF-teacher.SGM-ACC
saraha l-qasīdat-a
explain.PFV.3SGM DEF-poem.SGF-ACC
It seems that the teacher explained the poem. MSA: Salih (1985, 326)

b. *aṭ-ṭalib-at-u **ta-bdū** ʔanna-ha qad
DEF-student-SGF-NOM 3F-seem.IPFV.SG.INDIC that-3SGF.ACC QAD
qar-at al-kitāb
read.PFV-3SGF DEF-book
intended: The student (F) seems to have read the book.
*SSR MSA: Mohammad (2000)

Notwithstanding this evidence of the unavailability of SSR, Salih (1985) discusses structures of the type in (4), where the matrix clause involves the NP subject 'the girl', with the *seem* predicate displaying 3SGF agreement with it, which is also the SUBJ of the embedded predicate 'write'. The embedded verb 'write' shows 3SGF agreement, and the complementising element *kāʔanna* also shows 3SGF pronominal inflection.

(4) **bad-at-i** l-bint-u$_i$ **kaʔanna-hā**$_i$
seem.PFV-3SGF-INDIC DEF-girl-NOM as if-3SGF.ACC
katab-at-i r-risālat-a
write.PFV-3SGF-INDIC DEF-letter-ACC
The girl seemed as if she wrote the letter. MSA: Salih (1985, 138)

Camilleri et al. (2014) argue that structures such as (4) are in fact instances of **copy raising** constructions. In this case we are dealing with a SUBJ-to-SUBJ anaphoric dependency. As well as the anaphoric copy pronoun, in (4) we find the complementiser *kaʔanna*, onto which the SUBJ pronoun attaches (in ACC form). This complementiser is in a complementary distribution with *ʔanna* in MSA, and it appears to be that which enables the availability of a copy raising construction and its interpretation. As in English, it seems that the SUBJ in the matrix is understood as a PSOURCE, in such constructions.

Beyond MSA, ElSadek and Sadler (2015, 89-91) argue that the use of *kaʔinn*, the 'as if' complementiser in Egyptian Cairene Arabic, similarly results in the availability of a copy raising structure that involves an (individual) PSOURCE reading of the matrix SUBJ. The contrast in (5) is meant to demonstrate that while the use of the complementiser *ʔinn* (in this case with an attached pronoun), allows an expletive, non-raising, structure, the use of the complementiser *kaʔinn*, which obligatorily takes an attached pronoun, makes available a raising construction, and more precisely, a copy raising one. It is thus specifically the choice of the complementiser that determines a number of semantic and syntactic factors. This in turn suggests that in this construction, the functional category of C plays a key role in semantic interpretation.[1]

(5) a. bāyen (**ʔinn-ik**) mabsūt-a
show.ACT.PTCP.SGM that-2SG.ACC happy-SGF
It seems you (F) are happy.
it-expletive structure, Egyptian: ElSadek and Sadler (2015, 89)

b. mona$_i$ bāyn-a **kaʔinn-ha$_i$** mabsūt-a
mona show.ACT.PTCP-SGF as though-3SGF.ACC happy-SGF
Mona seems to be happy. CR, Egyptian: ElSadek and Sadler (2015, 90)

Examples such as those in (6), where the role of the SUBJ as PSOURCE is brought out rather clearly through the choice of the matrix predicate, provide further support for the idea that these sorts of examples in ECA are indeed copy raising structures. Observe once again the presence of the complementiser *kaʔinn*+pronoun, or its phonological variant *akin*+pronoun, which links the two clauses and hosts the copy raising pronominal itself.

(6) a. ḥassē-t akin-ni wiʔiʕ-ti min sābiʕ dōr
feel-PFV-1SG as though-1SG.ACC fall.PFV-1SG from seventh floor
I felt as though i fell from the seventh floor.

b. adī-ni wāʔif kaʔin-ni fil-ʔutubīs
still-1SG.ACC stand.ACT.PTCP.SGM as though-1SG.ACC in.DEF-bus
I'm still standing as though I am on the bus.
Egyptian: Woidich (1989, 124)

[1]Examples marked as Egyptian are Cairene Egyptian throughout.

2.3 Comparative (pseudo)-gap(s) in the Arabic raising system

To summarise, so far we have observed the following for Arabic, for the relevant class of matrix predicates:

1. expletive subject structures, with the *seem* predicate in default 3SGM form (or SGM form if the morphological nature of the predicate itself does not support PERS marking). If a complementiser is present at all it will be *ʔinn*;

2. copy raising constructions, in which the *seem* predicate occurs with a referential SUBJ and shows agreement with that subject. The choice of complementiser is *kaʔinn* and a pronominal copy coindexed with the matrix SUBJ occurs in the embedded predication.

Given this, there seems to be a distributional gap, with the absence of canonical SSR raising type constructions for this class of predicates. This does not mean that SSR is completely absent from the syntax of Arabic. Canonical SSR structures are widely available in the context of auxiliaries which may occur as c-structure functional categories and express a range of typically temporal, aspectual and modal type meanings. Phasal auxiliaries such as inceptive *badaʔa* 'start' or proximative *karaba* 'be about to', and others have been looked at descriptively in Mitchell and Ḥasan (1994), Maas (2009), Firanescu (2010), Saddour (2010), Naïm (2016), *interalia*), and have been analysed as raising predicates in Alotaibi et al. (2013), Wurmbrand and Haddad (2016), Camilleri (2016), and ElSadek (2016). Other auxiliaries which have also been analysed as having an f-structure PRED value with a non-thematic SUBJ and permitting a SSR construction are a range of pseudo-verbal sorts of auxiliaries which express aspectual as well as modal meanings. These auxiliaries are non-canonical in the sense that, while functioning as auxiliaries, they are usually themselves grammaticalisations out of Ps (hence the label 'pseudo-verb'). Examples of these include the pseudo-verbal auxiliary *il*+pronoun 'have' (Hallman (2016), Camilleri (2016), Camilleri and Sadler (2018)) (< P 'to'), and *baʕd*/*ʕad*+prn 'just, still' (< P 'after'), etc.

In fact the broad-brush observation above, that SSR constructions are limited to the more 'functional' (i.e. non-lexical) type of meanings, and are not found with predicates of perceptual report (such as *seem*), is not quite correct, and we will revise it further in §3.1 below. At this point, two further observations are in order. The first concerns the nature of the predicates used to express perceptual reports. In dialectal Arabic, canonical verbal predicates are **not** typical (although as we have illustrated above, such verbal predicates *do* occur in MSA). The second is that some of these non-verbal predicates in the vernaculars do in fact permit SSR. for Egyptian Cairene Arabic ElSadek and Sadler (2015) list the following:

1. ***bāyen*** 'show.ACT.PTCP.SGM' (< *bān* 'show, appear') +/- *ʕala*+Prn/NP 'on' - used in expletive and copy raising structures (as in (5a) and (5b), respectively);

2. The definite agent participle form: ***ʔiẓ-ẓāher*** 'the apparent', which is only used in expletive constructions. It is the derivationally-related counterpart of the raising predicate *yaḍhar* 'seem, appear' in MSA;

3. The pseudo-verbal form derived from the N ***šakl***+pronoun/NP lit. 'shape, form', available in SSR contexts with either individual or eventuality PSOURCE readings (as we will see in §3.1), as also illustrated for Jordanian in Jarrah and Alshamari (2017), and in copy raising constructions in the presence of *kaʔinn*+pronoun.

   To this list of non-canonical means, we can add:

4. The passive participle counterpart of the active participle *bāyen*: *mbāyen*, which can be used in the same way, along with the optional presence of the PP *ʕala*+Prn, as in (7)-(8).

(7) mā m-bāyen ʕlay-k el-kebr
NEG PASS.PTCP-appear.SGM on-2SGM.GEN DEF-oldness
You don't seem old. Lebanese: Feghali (1928, 6)

(8) m-beyyen (ʕeley-k) ǧōʕān / bidd-ak
PASS.PTCP-appear.SGM on-2SGM.GEN hungry.SGM / wish-2SGM.GEN
t-rūḥ
2-go.IPFV.SGM
You seem (lit. it appears (on you)) as though you (M) are hungry / as though you (M) want to go. Tulkarem Palestinian

Notwithstanding their generally restricted nature, canonical verb forms for perceptual reports which are the counterpart of MSA *yaḍhar* '3M-seem.IPFV.SG' *do exist* in some vernaculars. The data in (9), from Moroccan, Tunisian and Syrian, illustrate these usages, in which the verbal predicate appears to be constrained to expletive constructions, as in MSA.

(9) a. **ta-y-dher** belli kan-u hna
HABIT-3SGM-seem.IPFV that be.PFV.3-PL here
It seems that they were here. Moroccan: Benmamoun (2000, 125)

b. **yu-dhur** illi l-mṭar bāš t-ṣub
3M-appear.IPFV.SG that DEF-rain.SGF FUT 3F-pour.IPFV.SG
It seems that it is going to rain. Tunisian: Halila (1992, 243)

c. l-wlad **yə-ẓhar** ənnu ħak-u maʕ nawal
DEF-boy.PL 3M-seem.IPFV.SG that talk.PFV.3-PL with nawal
The boys, it seems that they talked to Nawal. Syrian: Farhat (1991, 164)

It is important to add that (9) are indeed true instances of *it*-expletive constructions, and that the use of the default 3SGM form of the verb here is a constraint associated with the construction itself, and not due to deficiencies in the morphological paradigm. Such evidence comes from the fact that *dher* in its other lexical (non-clause embedding) uses takes the usual full inflectional range, as illustrated through the inflected perfective 3PL (distinct) 'appear' forms in (10) below, based on data in Qwaider et al. (2018, 2):

(10) a. ʔāħla šiy ʔinnu aš-šabāba kill-uwn **ḍahar-uwā**
best thing that DEF-guy.PL all-3PLM appear.PFV-3PLM
ʕas-sāħih
on.DEF-scene
The best thing is that all the guys have appeared on the scene. Lebanese

b. aš-šabāb hallaʔa kullu-hum **dahar-uwā** ʕas-sāħah
DEF-guy.PL now all-3PLM.GEN appear.PFV-3PLM on.DEF-scene
All of the guys have now appeared on the scene. Jordanian

c. ʔāħla šiy ʔinnu aš-šabāb kill-uwn **bayan-uwā**
best thing that DEF-guy.PL all-3PLM appear.PFV-3PLM
ʕas-sāħih
on.DEF-scene
The best thing is that all the guys have appeared on the scene. Syrian

This difference is consistent with viewing the distinct use in an *it*-expletive, perceptual report construction as further along a grammaticalisation cline than the non-clause embedding use of this verb, in line with Kuteva et al. (2019, 10): 'decategorialization has the effect that the element concerned loses morphosyntactic properties characteristic of its less grammaticalized (e.g. lexical) source, such as the ability to take modifiers or inflections, and it shifts from a form class having many members (e.g. an open class) to one having only few members (a closed class).'

Even this is not the whole story as far as cross-dialectal microvariation is concerned. One of the ways in which Maltese expresses the meaning of 'seem, appear' is with the verb *deher* (*jidher* in the imperfective), the counterpart of the verb employed in (9) for other dialects, and *yaḍhar* in MSA. As the data in (11) shows, the lexical verb in Maltese allows for all of the available constructions, *it*-expletive, SSR and copy raising structures, which can in turn all appear with or without the complementiser *li* 'that'.

(11) a. **j-i-dher** (li) (it-tfal) sejr-in
3M-FRM.WVL-appear.IPFV.SG that DEF-children go.ACT.PTCP-PL
tajjeb
good
It seems that the children are doing well. *it*-expletive structure

b. (it-tfal) **j-i-dhr-u** (li) sejr-in
DEF-children 3-FRM.VWL-appear.IPFV-PL that go.ACT.PTCP-PL
tajjeb
good.SGM
the children seem to be doing well. SSR: Camilleri (2018, 172)

c. **t-i-dher** (li) ġà ta-w-**ha**
3F-FRM.VWL-appear.IPFV.SG that already give.PFV.3-PL-3SGF.ACC
xebgħa xogħol x't-a-għmel
smacking work what.3F-FRM.VWL-do.IPFV.SG
she seems as though they already gave her a lot of work to do. CR: Camilleri et al. (2014, 192)

While the *verbal* predicate corresponding to *yaḍhar* in MSA permits only the expletive construction, the synchronic situation across the Algerian dialects appears to be that they employ the verb-form *bēn* 'seem, appear', which is the verb associated with the active and passive participles *bāyen* (5) and *mbāyen* (7), in all three constructions, i.e. expletive, SSR and copy raising structures, as in (12).

(12) a. **y-bēn** billi/ʔinnu štī-t-u
3M-appear.IPFV.SG that love.PFV-3SGF-3SGM.ACC
It seems that she loved him. *it*-expletive structure

b. **kun-t-i** **t-bēn-i** ∼ **bēyn-a**
be.PFV-2-SGF 2-appear.IPFV-SGF ∼ appear.ACT.PTCP-SGF
ti-bɣ-i t-ruħ-i / rāki ʕayyān-a
2-want.IPFV-SGF 2-go.IPFV-SGF / COP.2SGF tired-SGF
You seemed to want to go. / You seemed (to be) tired. SSR

c. **t-bēn-i** **killi** rāki ʕayyān-a
2-appear.IPFV-SGF as though COP.2SGF tired-SGF
You seem like you're tired. SSR oblig. individual PSOURCE reading

d. **t-bēn-i**$_i$ **killi** darb-u-**ki**$_i$
2-appear.IPFV-SGF as though hit.PFV.3-PL-2SGF.ACC
You seem like they've hit you. CR: S. Rouabah PC

What has not been previously observed (to our knowledge) is that an individual PSOURCE reading necessarily arises, whether in a SSR structure (as in (12c)) or a copy raising structure (as in (12d)) in the presence of the complementiser *killi*, which is both the structural and semantic counterpart to *kaʔanna* in other dialects.

Beyond the use of these 'seem' verbal predicates across the different dialects, South Western Saudi dialects have grammaticalised their own idiosyncratic verbal expression of perceptual reports. In this case, the entire semantics of the verb *talaʕ*, which otherwise means ‘go up’ (and even ‘go out’ at times) has been abandoned in this additional lexical meaning of ‘appear’, together with the development of a SSR syntax alongside an *it*-expletive alternative. The contrastive lexical uses are provided in (13) below:

(13) a. ṭalaʕ-at f’ɣurfat-ha
go up.PFV-3SGF in.room.SGF-3SGF.GEN
She went up to her room. <SUBJ, OBL>

b. ti-ṭlaʕa m-sāfr-a
3F-go up.IPFV.SG ACT.PTCP-abroad-SGF
She appears to be abroad (from some sort of inference). SSR

c. ṭalaʕ innu hiya m-sāfr-a
go up.PFV.3SGM that 3SGF.NOM ACT.PTCP-abroad-SGF
It seemed that she is abroad (as she was not answering my calls, for instance). *it*-expletive <COMP>SUBJ

There is textual evidence in Palestinian (also confirmed by native speakers) of the use of *ṭalaʕ* in a non-clause embedding ‘appear’ sense in (14) (very much in parallel to the verbs *bēn* and *dher* in their pure ‘appear’ uses given in (10)). Uses such as this in which ‘appear’ occurs with a PP oblique may well have been the bridge allowing for the emergence of structures such as (13b-c), via changes to the a-structure (from OBL to XCOMP/COMP, giving rise to the SSR and *it*-expletive constructions in (13b-c)).

(14) aš-šabāb kulla-hum **ṭalʕ-uwā** ʕas-sāħah
DEF-guy.PL all-3PLM.GEN appear.PFV-3PLM on.DEF-scene
All the guys have appeared on the scene. Palestinian: Qwaider et al. (2018, 2)

3 Grammaticalisation

We can conclude from the overview of data in §2.3 that there are limited instances of canonical verbal predicates of perceptual report which occur in raising structures. In what follows we extend the discussion of what we might call non-canonical strategies for expressing perceptual reports involving raising structures in Cairene Arabic in ElSadek and Sadler (2015) to a range of data from other dialects. We propose two grammaticalisation paths, which we argue have compensated for the absence of canonical lexical raising predicates and SSR constructions. In earlier work, Barron (1997); Barron (2001) provides a diachronic account of raising predicates within an LFG framework. She considers the grammaticalisation of perceptual verbs, in particular physical visual perception verbs. The grammaticalisation that results in the formation of the raising predicates in the cases Barron discusses differs most notably in that it involves at the onset verbal predicates with a more elaborate argument-structure, whereby it is as a result of a suppression of the perceiver argument that a distinct argument-structure - functional-structural mapping results. Coupled with the presence of secondary predication and bleached semantics, a new meaning is lexicalised, and a raising construction grammaticalised.

The paths we choose to consider here involve the formation of new lexical items, which we take to be synchronically verbal although their source is clearly in other lexical categories. Of course they can be completely verbal in terms of f-structure subcategorisation and still maintain categorial and morphological vestiges of the diachronic source. Moreover, they don't have to be at the same stage in every dialect, with transitions further along a grammaticalisation path in some, rather than in others. In the rest of this paper we discuss two paths which we argue have in turn resulted in the development of a clausal (raising) predicate of perceptual report from a N and a P respectively.

Before delving into the individual paths, we make reference to the fact that as commonly present in the context of grammaticalisation trajectories, the N under analysis is concurrently maintained in the system as a canonically-behaving lexical N, over and above the distinct use it has developed through time. The result is an instance of a **functional split**, whereby the grammaticalisation path of change undertaken by this particular lexical item has not resulted in the item's loss. Rather, the effect is such that a second function complements its existing use in the grammar, giving a *layering effect* (Hopper and Traugott, 2003). This is also true of the use of *dher* in dialects such as Jordanian and Lebanese, illustrated in (10), where we seem to have alternate argument-structures (<SUBJ,(OBL)> vs. <XCOMP>SUBJ) with accompanying morphosyntactic differences. The same does not hold of the P under analysis. It is not anymore in use as a lexical preposition, in the variety we will be looking at.

3.1 *šakl*+prn > raising lexical V

The use of the N *šakl* lit. 'shape, form' as a lexical raising predicate embedding a clausal argument, described and analysed for Egyptian in ElSadek and Sadler (2015) is commonly found in the dialects from Libya eastwards. When used with this function, the predicate takes on the meaning of 'seem, appear'. We start by establishing the lexical behaviour of *šakl* as a noun.

(15) a. šakl id-daerah mdawwar/mdawwar-ah
shape.SGM DEF-circle.SGF round.SGM/round-SGF
The shape of the circle is round.

b. id-daerah, šakl-aha mdawwar/mdawwar-ah
DEF-circle.SGF, shape.SGM-3SGF.GEN round.SGM/round-SGF
The circle, its shape is round.

c. ana miš ʕağib-nī iš-šakl il-kbīr
I NEG please.PFV.3SGM-1SG.ACC DEF-shape.SGM DEF-big.SGM
la-l-bayt
to/for-DEF-house.SGM
I do not like the big shape of the house. Palestinian: Al-labadi, PC

Notice in (15) the alternative agreement forms possible on the predicative adjective 'round'. The syntactic head of the SUBJ is the N *šakl* 'shape, form' which is SGM, and the grammatically expected form for a predicative adjective agreeing with a SGM subject would be SGM. The occurrence of SGF indexing on the predicative adjective suggests that in this case, it is agreeing with the dependent argument within the NP SUBJ, *id-daerah* 'the circle', which is SGF, and which together with the head *šakl* forms a construct state construction. The f-structure corresponding to (15a) is shown in (16), where we represent the dependent argument as a POSS GF subcategorised by the head noun in SUBJ GF. We suggest that the availability of this (otherwise unexpected) agreement pattern is suggestive of **prominent internal possessor** behaviour (see Nikolaeva et al. (2019), and more below), and we argue is key to the reanalysis which underlies the development of this lexical item into a raising predicate.

(16)
$$\begin{bmatrix} \text{PRED} & \text{`MDAWWARAH}\langle\text{SUBJ}\rangle\text{'} \\ \text{TENSE} & \text{PRESENT} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`ŠAKL}\langle\text{POSS}\rangle\text{'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \\ \text{POSS} & \begin{bmatrix} \text{PRED} & \text{`ID-DAERAH'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \\ \text{DEF} & + \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

We now turn to examples illustrating the use of *šakl* as something other than a simple noun. In an example such as (17), *šakl* is not a dependent nominal argument, but is the matrix predicate of the sentence, occurring with a SUBJ (*Morsi*) and a clausal complement. The *form* of (non-canonical) agreement with the SUBJ (which is glossed here as 3SGM.GEN) reflects the nominal origin of this pseudo-verbal predicate.

(17) Morsi šakl-u rigiʕ
Morsi shape-3SGM.GEN return.PFV.3SGM
Morsi seems to have come back. Egyptian: ElSadek and Sadler (2015, 96)

As argued in ElSadek and Sadler (2015), pseudo-verbal *šakl* is a raising predicate. In (18) it occurs with a (raised) non-thematic weather-verb SUBJ, as seen by the presence of the pleonastic 3SGF form selected by weather predicates. This in turn suggests that the bound GEN pronoun on *šakl* can no longer be associated with the POSS GF it marks in nominal examples such as (15).

(18) šikil-**ha** ʔib-ti-šti
shape-3SGF.GEN B-3F-rain.IPFV.SG
It seems to be raining. Jordanian: Jarrah and Alshamari (2017, 33)

In the SSR construction, the raising predicate *šakl* allows both individual and an eventuality PSOURCE readings of the matrix SUBJ. In the actual context of utterance for (17), *Morsi* was in fact dead, and hence the perceptual source is the eventuality, not the individual, and in (19), as illustrated by the authors, the context involves an inference from a phone conversation, and thus not a direct perception of the individual in question.

(19) šikil-ak ma ħaḍḍar-it-š malīh
shape-2SGM.GEN NEG prepare.PFV-2SGM-NEG well
You seem to not have prepared well (for the exam). (Inferred from a description during a phone conversation).
Jordanian: Jarrah and Alshamari (2017, 32)

An individual PSOURCE reading, on the other hand, is possible with both SSR and copy raising constructions which have *šakl*+pronoun as a matrix raising lexical predicate, and in fact an individual PSOURCE SUBJ is obligatory in the latter structure.

(20) a. kān/kon-t-i šakl-ik
be.PFV.3SGM/be.PFV-2-SGF shape-2SGF.GEN
bi-t-ħib-ī-h
B-2-love.IPFV-SGF-3SGM.ACC
You seemed to love him. SSR, Egyptian: ElSadek and Sadler (2015, 97)

b. šakl-ak mrīð
shape-2SGM.GEN sick.SGM
You seem sick. SSR, Benghazi Libyan: Saad (2019)

c. šakl-**ak** kaʔinn-**ak** mabsūt
shape-2SGM.GEN as though-2SGM.ACC happy.SGM
You seem as if you are happy. CR, Egyptian: ElSadek and Sadler (2015, 99)

d. šakl-**aha**$_i$ **kaʔinnu-hum**$_j$ deħk-u$_j$ ʕalē-**ha**$_i$
shape-3SGF.GEN as though-3PL.ACC laugh.PFV.3-PL on-3SGF.GEN
She seems as if they've fooled her. CR, Egyptian: ElSadek and Sadler (2015, 99)

The synchronic end result is the formation of a **pseudo-verbal form**, which in the literature on Arabic (Ingham (1994), Vanhove (1993), Brustad (2000), Comrie (2008), Ingham (2008), Peterson (2009), and Vanhove et al. (2009)) refers to a class of forms, be they prepositions, nouns, quantifiers, etc. that have taken on a verb-like function, and express the reanalysed SUBJ GF via non-canonical inflectional forms, since these items maintain their erstwhile GEN pronominal forms/inflection. As well as having very different semantics and functions, the nominal and pseudo-verbal forms participate synchronically in distinct structures, with the development of the 'seem, appear' meaning from the original 'shape, form' also resulting in the availability of a new, raising construction. The f-structure associated with (21), repeated from (18) above is provided below.

(21) šikil-**ha** ʔib-ti-šti
shape-3SGF.GEN B-3F-rain.IPFV.SG
It seems to be raining. SSR - Jordanian: Jarrah and Alshamari (2017, 33)

$$\left[\begin{array}{lll} \text{PRED} & \multicolumn{2}{l}{\text{`ŠIKIL}<\text{XCOMP}>\text{SUBJ'}} \\ \text{TENSE} & \multicolumn{2}{l}{\text{PRESENT}} \\ \text{SUBJ} & [1] & \left[\begin{array}{ll} \text{PRED} & \text{`PRO'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{array}\right] \\ \text{XCOMP} & & \left[\begin{array}{ll} \text{PRED} & \text{`ʔIBTIŠTI}<\text{SUBJ}>\text{'} \\ \text{SUBJ} & [1] \end{array}\right] \end{array}\right]$$

3.2 Diachronic Trajectory

We hypothesise that the development of *šakl* as used in a structure such as that in (15a), where its semantic form is '*šakl*<POSS>', to one where it functions as a raising predicate, as in (21), with the semantic form '*šakl*<XCOMP>SUBJ' progressed primarily out of a predicative construction of the type in (15a), where the adnominal possessor appears to exhibit properties that trigger behaviours associated with prominent internal possessors (PIP). For Arabic, there is clear and uncontroversial evidence that these nominal examples and other adnominal/NP-internal possessive constructions in general involve an internal possessor/distinguished argument. In Arabic, the possessor and the possessed together form a tightly-knit and inseparable morphosyntactic unit, often called a construct state construction (after the form of the head noun, which occurs in a particular form if SGF) or annexation structure. Although not a 'direct' argument of the PRED, as it is not selected by the construction's predicate but rather by the predicate *šakl* in SUBJ position, the internal possessor nevertheless may optionally trigger agreement/indexing on the clausal predicate, which in this case was *mdawwar* 'round', as if it were a direct dependent of this predicate. The hypothesis we suggest here is that the status of the internal possessor as a PIP is an enabling or triggering factor for the development of a verbal raising predicate out of the N *šakl*. Nikolaeva et al.'s (2019) typological investigation demonstrates that PIPs 'are most likely to stand in an alienable relation with the possessed noun' (p. 26), and this is precisely the case with *šakl* vis-à-vis the possessor. Nikolaeva et al. (2019, 8) discuss possibilities as to how such non-canonical agreement effects with prominent possessors can be resolved synchronically, including the entertaining of syntactic loosening of the notion of locality, as well as the functional prominence of PIPs, which they consider to be partly semantic and partly associated with information-structure associated prominence effects (p. 24). Referential features of the possessor, including its salience with respect to prominence and the Animacy Hierarchy, as well as its affectedness and involvement in the event, all contribute to the likelihood of PIP behaviour. An LFG treatment which combines both syntactic and information-structure mechanisms in accounting for the properties of PIPs, such as serving as agreement controllers, is developed in Ritchie (2016). In this account, the PIP essentially also serves as a secondary topic, and hence, as an agreement controller (Dalrymple and Nikolaeva, 2011). It is not our task here to discuss the motivations for the PIP properties present in structures such as (15a), or how they should be accounted for or motivated, but we take it to have been a contributory factor at the **diachronic onset** in the development of a SSR across a number of Arabic vernaculars. We argue that as the possessor θ-role stands higher (in terms of saliency) than the possessed counterpart, the θ-role - GF mapping displays something distinct, with the lower θ-role being mapped onto a more prominent GF. This in turn creates a tension, giving rise to PIP-associated behaviours for the POSS, superseding the most prominent GF in these aspects. We believe this to be true especially in contexts where the possessor is expressed as a pronominal form (which is in turn higher on the Animacy scale than a NP), as in fact this is very much the key property of the use of *šakl* in the raising structures that have developed. Gradually these PIP behaviours grammaticalise, at least in certain contexts (structural or semantic), such that the possessor starts being reinterpreted as the SUBJ. This in turn causes ripples not merely to the morphosyntactic dimension of the structure, where the bound GEN pronominal form on *šakl* is reanalysed as a non-canonical SUBJ exponent, but additionally to the organisation of the entire f-structure itself (even if the c-structure may remain totally unchanged). Given the reinterpretation of the GEN pronominal form as the SUBJ exponent, which is otherwise a behaviour attributive of verbal PREDs, as these are the categories that canonically allow for bound pronominal SUBJ GFs, the N *šakl* takes on a pseudo-verbal function, which is a common place grammaticalisation and lexicalisation process across the Arabic macrosystem, where it ends up as the structure's highest PRED as some sort of V, to which the original structure's PRED, i.e. the predicative adjective becomes subordinated. This newly formed pseudo-verb now not only takes the SUBJ GF in its scope, which was the original POSS, but it additionally takes the former PRED as an argument, reanalysed as an XCOMP).[2] An open clausal argument must be assumed to allow for the persistence of the dependency that existed between the original predicative adjective (now the reanalysed XCOMP) and the SUBJ or the SUBJ POSS. Once the XCOMP GF established itself, this then paved open the way for that clausal argument's mapping onto different constituents at the c-structure, following which, changes at the c-structure are observed. As noted by an external reviewer, whom we thank for sharing with us this insight, while the grammaticalisation just exemplified above involves a striking development entailing a **categorial shift** from a N to a V, the **semantic trajectory** finds parallels in other languages. One such parallel is the predicate *semblar* 'seem' in French, which in turn is derived from the Latin verb *simulare* 'copy, pretend', which is itself related to the noun *simulacrum* 'shape, form, copy'.

3.3 P 'like' > raising lexical V

The P that seems to have led to the general evolution of raising, not solely in the context of the SSR we are exploring here, which is *specific* to a particular vernacular, but also to the copy raising constructions found in the Arabic macrosystem (including MSA) is the P 'like, as', as demonstrated rather extensively and in detail, in Taine-Cheikh's (2004) descriptive study of this element. Here we argue that the grammaticalisation of this P leads to the formation of raising structures from this source. Moreoever, although we do not make this argument in detail here, part of the picture motivating and supporting our view is the fact that the CR structures which we illustrated in 2.2 involve the complementisers *kaʔanna* or *killi* (depending on the variety - see for example the Algerian copy raising constructions in (12c-d)) which are themselves the fusion of two items, the P *ka*, *kif* 'like' + the complementiser *ʔann* or *(il)li*, in a diachronic process of **univerbation**.[3] In this respect, this aligns Arabic very much with typologically unrelated copy raising structures such as those found in English, which in turn obligatorily require the use of *like* or *as* as mediating Ps between the matrix and embedded clause.

Here we present data in which (another) P 'like' appears to have resulted in the grammaticalisation and development of a SSR construction. Such unconventional, yet straightforward examples which suggest the development of a P 'like' into a lexical raising predicate are the pseudo-verbal uses of *zēy* 'like' + pronoun in

[2]This development we can take to constitute the instantiation/formation of a secondary predication, which according to Barron (1997); Barron (2001) is one of the conditions that allows for perception verbs to become raising predicates.

[3]The fusion that produced the complementiser *billi* 'that', illustrated in (9a) and (12a), for Moroccan and Algerian, respectively, follows the same sort of diachronic process with a different preposition.

Algerian, which are translated in the original text as: '*il paraîtrait que*, *il a l'air de*, *il semble*'. The structures below *might* be analysed as SSR constructions; in which case, the f-structure associated with (22b) would be as shown in (23).

(22) a. **zēy-u** nsā-na
like-3SGM.GEN forget.PFV.3SGM-1PL.ACC
He seems to have forgotten us.

b. **zēy-ik** ku-te-t-mesḥor b-en-nās
like-2SG.GEN HABIT-2SG-REFL-mock.IPFV with-DEF-people
You seem to be making fun of people.

c. **zēy-na** mberrd-în
like-1PL.GEN cold-PL
We seem cold. Djidjelli Algerian: Marçais (1954, 524)

(23)
$$\begin{bmatrix} \text{PRED} & \text{`Z}\bar{\text{E}}\text{Y<XCOMP>SUBJ'} \\ \text{TENSE} & \text{PRESENT} \\ \text{SUBJ} & [1]\begin{bmatrix} \text{PRED} & \text{`PRO'} \\ \text{PERS} & 2 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`KUTETMESḤOR<SUBJ, OBL>'} \\ \text{ASPECT} & \text{HABITUAL} \\ \text{SUBJ} & [1] \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`BI<OBJ>'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`N}\bar{\text{A}}\text{S'} \\ \text{DEF} & + \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Given the lack of additional data, this is necessarily somewhat speculative and while there is evidence for a control relation, we do not have evidence that bears directly on the question of whether the SUBJ is thematic or non-thematic in the *zēy* clause, and hence, these could instead be instances of equi-type structures with obligatory SUBJ control.

The pseudo-verbal strategy using *zēy* (originally a P, and still functioning as such in other dialects), which we suggest may be another instance of a pseudo-verbal SSR construction, is synchronically a receding, if not a completely archaic strategy. Young Algerian speakers find the use of *zēy* rather archaic, or only associated with Egyptian market sellers (personal communication, Algerian colleagues). The synchronic P meaning 'like' is *kif*, which as we have seen above, also occurs fused with the declarative complementiser to mark copy raising constructions. Nevertheless, some further evidence of the P *zēy* is available for distinct dialects of Algerian, as it is found in Bedouin Algerian dialects documented in 1908.

(24) lābes zēy el-mɣāṛba
wear.ACT.PTCP.SGM like DEF-Moroccan.PL
He was wearing (i.e. dressed) like Moroccans. Saïda Algerian: Marçais (1908, 175)

Synchronically, across the different Algerian dialects, it is the lexical raising verbal predicate *bēn* 'appear, seem' that is used instead of *zēy* for perceptual reports, as illustrated above in (12).

We can offer only some highly speculative remarks concerning possible grammaticalisation paths from the P 'like' to a perceptual report predicate heading a SSR construction, with concomitant changes in lexical meaning and argument-structure as *zēy* shifted from a P to a V. We take it that at the outset we have a non-verbal predicative construction headed by the P *zēy* 'like'. In uses corresponding to *He is like me.* the P would have a PRED '*zēy*<SUBJ, OBJ>'. It is possible that (as a reviewer suggests) the P predicate might have also allowed a more abstract sense in which the non-subject argument is a COMP, along the lines of *It/the situation is like/as if we will be leaving*, with the closed COMP developing over time into an open XCOMP with subject re-entrancy. Two further issues would arise in relation to such a trajectory. The first is that elsewhere in Arabic (i.e. with other predicates), we seem to find rather the reverse development, in as much as we find instances where a predicate has (additionally) developed COMP uses (with default 3SGM agreement on the predicate) after the establishment of XCOMP uses. The second is that a question arises as to why/how what would be the SUBJ of the COMP is realised as though it were the OBJ of the P *zēy*, which is then in turn reanalysed leading to the emergence of a quite prototypical case of a pseudo-verb with the attached pronoun reanalysed as the (non-canonical) exponent of the SUBJ of *zēy* itself. Given the lack of a historical record and our current state of knowledge, we cannot be more concrete at this stage.

4 Conclusion

In this paper we have looked at the development of non-canonical predicates of perceptual report from two distinct non-verbal lexical items, a N and a P, which have given rise to subject-to-subject raising constructions. Diachronically, the processes which have given rise to the development of these verbal raising predicates do not involve the emergence of a functional (featural) meaning through loss of content and bleaching, although this is something which is common place, and the normal domain of grammaticalisation. Rather, what can be observed in each case is a semantic development in which a lexical shift from a concrete to a more abstract sense is involved, in parallel to the observation Barron (1997, 12); Barron (2001, 73) makes with respect to physical visual perception verbs in a number of languages as they grammaticalise into raising predicates, where what we find is a 'cognitive shift from a physical to a mental process'. In the case of the vernacular

Arabic data set presented here, the concrete > abstract sense shift is coupled with a concomitant emergence of a more elaborate argument- and functional-structure. This is distinct from the diachronic account provided by Barron. In the data she discusses, the change that results in a raising predicate primarily involves predicates that take a full argument-structure themselves, and which over time undergo a reanalysis triggered via a suppression of one of the relevant arguments. In each of the two case studies considered here, the semantic and functional changes that take place are what lead to the creation of a construction involving a N or a P that becomes solidified and fixed in its meaning to such an extent that it begins to get reanalysed as encoding perceptual reports– a meaning which is then invested in the lexical nexus of that construction. In the case of the nominal source, *šakl* 'shape, form', we suggest that a key role is played in these diachronic developments by the presence of a prominent internal possessor. The result is an uncommon categorial shift, but which displays a semantic trajectory that is found elsewhere. The same is holds in the case of the development of the P, whereby while the path from P to V might be unusual, the development of a perceptual report predicate from 'like' is semantically very natural, and consistent with other developments in Arabic and beyond, notably in marking copy raising constructions themselves.

References

Alotaibi, Yasir, Muhammad Alzaidi, Maris Camilleri, Shaimaa ElSadek, and Louisa Sadler. 2013. Psychological Predicates and Verbal Complementation in Arabic. In M. Butt and T. H. King, eds., *Proceedings of LFG13*, pages 6–26. Stanford, CA: CSLI.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Asudeh, Ash and Ida Toivonen. 2012. Copy raising and Perception. *Natural Language and Linguistic Theory* 30/2:321–380.

Barron, Julia. 1997. LFG and the history of raising verbs. In M. Butt and T. H. King, eds., *Proceedings of LFG97*, pages 1–13. Stanford: CSLI Publications.

Barron, Julia. 2001. Perception and raising verbs: synchronic and diachronic relationships. In M. Butt and Tracy Holloway King, eds., *Time over Matter: Diachronic perspectives on morphosyntax*, pages 73–103. Stanford: CSLI Publications.

Benmamoun, Elabbas. 2000. *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford: Oxford University Press.

Brustad, Kristen. 2000. *The Syntax of Spoken Arabic*. Washington, DC: Georgetown University of Press.

Camilleri, Maris. 2016. *Temporal and aspectual auxiliaries in Maltese*. Ph.D. thesis, University of Essex.

Camilleri, Maris. 2018. On raising and copy raising in Maltese. In P. Paggio and A. Gatt, eds., *The Languages of Malta*, pages 171–201. Berlin: Language Science Press.

Camilleri, Maris, Shaimaa ElSadek, and Louisa Sadler. 2014. Perceptual Reports in (varieties of) Arabic. In M. Butt and T. H. King, eds., *Proceedings of LFG14*, pages 179–199. Stanford, CA: CSLI Publications.

Camilleri, Maris and Louisa Sadler. 2018. Schematising (morpho)syntactic change in LFG: Approaches from grammaticalisation in Arabic. In Miriam Butt and Tracy Holloway

King, eds., *Proceedings of LFG18*, pages 108–128. Stanford: CSLI Publications.

Comrie, Bernard. 2008. Pseudoverb. In *Encyclopaedia of Arabic Language and Linguistics*, vol. III, pages 739–740. Leiden: Brill.

Dalrymple, Mary and Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

ElSadek, Shaimaa. 2016. *Verbal Complementation in Egyptian Colloquial Arabic: An LFG account*. Ph.D. thesis, University of Essex.

ElSadek, Shaimaa and Louisa Sadler. 2015. Egyptian Arabic perceptual reports. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of LFG15*, pages 84–102. Stanford : CSLI Publications.

Farhat, Haytham. 1991. *Empty categories in Syrian Arabic*. Ph.D. thesis, University of Wales, Bangor.

Feghali, Michel T. 1928. *Syntaxe des parlers arabes actuels du Liban*. Paris: Paul Geuthner.

Firanescu, Daniela R. 2010. Do you still love Feiruz? The modal bəʼa in spoken Arabic from Syria. *Monde Arabe* 7:123–142.

Halila, Hafedh. 1992. *Subject specificity effects in Tunisian Arabic*. Ph.D. thesis, University of Southern California.

Hallman, Peter. 2016. The universal perfect in Syrian Arabic. *Brill's Annual of Afroasiatic Languages and Linguistics* 8(1):76–100.

Hopper, Paul J. and Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Ingham, Bruce. 1994. *Najdi Arabic: Central Arabian*. Amsterdam/Philadelphia: John Banjamins.

Ingham, Bruce. 2008. Najdi arabic. In K. Versteegh, ed., *Encyclopaedia of Arabic Language and Linguistics*, vol. II, pages 328–334. Leiden: Brill.

Jarrah, Marwan and Murdhy R Alshamari. 2017. The syntax of the evidential particle *šikil* in Jordanian Arabic. *Italian Journal of Linguistics* 29(2):29–56.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Seongha Rhee, and Heiko Narrog. 2019. *World lexicon of grammaticalization*. Cambridge University Press.

Maas, Utz. 2009. Complex predicates in Maltese: From a neo-Arabic perspective. In B. Comrie, R. Fabri, E. Hume, M. Mifsud, T. Stolz, and M. Vanhove, eds., *Introducing Maltese Linguistics*, pages 113–132. Amsterdam/Philadelphia: John Benjamins.

Maling, Joan. 1983. Transitive adjectives: A case of categorial reanalysis. In F. Heny and B. Richards, eds., *Linguistic categories: Auxiliaries and related puzzles*, pages 253–289. Dordrecht: Reidel Publishing Company.

Marçais, Philippe. 1954. *Textes arabes de Djidjelli*. Paris: Presses universitaires de France.

Marçais, William. 1908. *Le dialecte arabe des Ulâd Brāhîm de Saïda (Département d'Oran)*. Paris: H. Champion.

Mitchell, Terence F and Shāhir Ḥasan. 1994. *Modality, Mood, and Aspect in Spoken Arabic: With Special Reference to Egypt and the Levant*. New York: Routledge.

Mohammad, Mohammad A. 2000. *Word order, agreement and pronominalization in Standard and Palestinian Arabic*, vol. 181. Amsterdam: John Benjamins.

Naïm, Samia. 2016. On the interaction between external and internal markers in expressing aspect in Arabic dialect varieties. In Z. Guentcheva, ed., *Aspectuality and Temporality: Empirical and theoretical issues*, pages 325–354. Amsterdam/Philadelphia: John Benjamins.

Nikolaeva, Irina, András Bárány, and Oliver Bond. 2019. Towards a typology of prominent internal possessors. In A. Bárány, O. Bond, and I. Nikolaeva, eds., *Prominent Internal*

Possessors, pages 1–38. Oxford: Oxford University Press.

Peterson, John. 2009. "Pseudo-verbs": An analysis of non-verbal (co-)predication in Maltese. In B. Comrie, R. Fabri, E. Hume, M. Mifsud, T. Stolz, and M. Vanhove, eds., *Introducing Maltese Linguistics*, pages 181–205. Amsterdam/Philadelphia: John Benjamins.

Qwaider, Chatrine, Motaz Saad, Stergios Chatzikyriakidis, and Simon Dobnik. 2018. Shami: A Corpus of Levantine Arabic Dialects. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Ritchie, Sandy. 2016. Two cases of prominent internal possessor constructions. In D. Arnold, M. Butt, B. Crysmann, T. H. King, and Stefan Müller, eds., *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, pages 620–40. Stanford: CSLI Publications.

Saad, Aisha. 2019. L'évolution sémantique de *šōr* (vers, en direction de) dans le dialecte de Benghazi (Libye). Poster presented at NACAL47, Paris.

Saddour, Inès. 2010. *Relating two simultaneous events in discourse: The role of ongoingness devices in L1 Tunisian Arabic, L1 French and L2 French by Tunisian learners*. Ph.D. thesis, Aston University.

Salih, Mahmud. 1985. *Aspects of Clause Structure in Standard Arabic: A Study in Relational Grammar*. Ph.D. thesis, SUNY at Buffalo.

Soltan, Usama. 2007. *On Formal Feature Licensing in Minimalism: Aspects of Standard Arabic Morphosyntax*. Ph.D. thesis, University of Maryland.

Taine-Cheikh, Catherine. 2004. De la grammaticalisation de 'comme'(comparatif) en arabe. In M. Haak, R. de Jong, and K. Versteegh, eds., *Approaches to Arabic Dialects. Collection of articles presented to Manfred Woidich on the Occasion of his Sixtieth Birthday*, pages 309–328. Leiden: Brill.

Vanhove, M. 1993. *La langue maltaise: Etudes syntaxiques d'un dialecte arabe "périphérique"*. Wiesbaden: Harrassowitz.

Vanhove, Martine, Catherine Miller, and Dominique Caubet. 2009. The grammaticalisations of modal auxiliaries in Maltese and Arabic vernaculars of the Mediterranean area. In J. can der Auwera, ed., *Grammaticalization of Modal particles*, pages 325–362. Berlin: Mouton de Gruyter.

Woidich, Manfred. 1989. Illi "dass", illi "weil" und zayy illi "als ob": Zur reinterpretation von relativsatzgefügen im kairenischen. *Mediterranean Language Review* 4:109–128.

Wurmbrand, Susi and Youssef A. Haddad. 2016. Cyclic spell-out derived agreement in Arabic raising constructions. In Youssef A. Haddad and Eric Potsdam, eds., *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics, Gainesville, Florida 2014*, pages 193–226. Amsterdam/Philadelphia: John Benjamins.

Constraining Scope Ambiguity in LFG+Glue

Matthew Gotham
University of Oxford

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 111–129

http://csli-publications.stanford.edu/LFG/2019

Keywords: scope, quantification, Glue semantics

Gotham, Matthew. 2019. Constraining Scope Ambiguity in LFG+Glue. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 111–129. Stanford, CA: CSLI Publications.

Abstract

A major strength of the Glue approach to semantic composition for LFG is that it accounts for quantifier scope ambiguity without the need for additional assumptions. However, quantifier scope is more rigid in some languages and constructions than Glue would lead us to expect. I propose a mechanism for constraining scope ambiguity in LFG+Glue, based on ideas taken from Abstract Categorial Grammar. Unlike existing proposals, this account does not depend on representational constraints on linear logic derivations or meaning representations.

1 Introduction

1.1 Scope ambiguity

Famously, sentences like (1) are ambiguous in English between a 'surface scope' and an 'inverse scope' interpretation.

(1) A police officer guards every exit.

The surface scope interpretation can be paraphrased as 'there is a police officer who guards every exit', and is represented logically in (2-a). The inverse scope interpretation can be paraphrased as 'every exit is guarded by a police officer', and is represented logically in (2-b).

(2) a. $\exists x.\text{officer}'x \wedge \forall y.\text{exit}'y \rightarrow \text{guard}'xy$
b. $\forall y.\text{exit}'y \rightarrow \exists x.\text{officer}'x \wedge \text{guard}'xy$

Pre-theoretically, we can refer to the ambiguity of (1) under consideration as 'scope ambiguity'. The surface scope interpretation is so called because the order of quantificational expressions on the surface matches the order of quantifers in the interpretation, and *mutatis mutandis* for the inverse scope interpretation. A question that immediately arises is whether or not this ambiguity corresponds to any distinction in possible syntactic representations or derivations of (1), or, more generally and succinctly:

Q Is quantifier scope ambiguity syntactic ambiguity?

Montague (1973) famously answered 'yes' to **Q**, as indeed the categorial grammar perspective forces one to. Cooper (1983) answered 'no', and explicitly criticized Montague on those grounds. Since May (1985), mainstream Chomskyan syntacticians and semanticists have tended to side with Montague (on this issue!), viewing (1) as having two different representations at a level of syntax known as

†Thanks to Mary Dalrymple, Jamie Findlay, John Payne, Adam Przepiórkowski and the audiences at LFG'19 and SE-LFG28 for helpful comments and discussion. This research is supported by an Early Career Fellowship from the Leverhulme Trust.

Logical Form (LF), each corresponding to one of the interpretations in (2). But note that the question of the existence of LF, and the connected debate over 'direct compositionality' (Jacobson, 2002; Barker & Jacobson, 2007), is not the same as **Q**. The systems for handling quantifier scope ambiguity outlined by Barker & Shan (2014) and Jacobson (2014) are both directly compositional, and yet they postulate a difference in syntactic derivation to account for the two different interpretations of (1).

Whenever **Q** has been asked in an LFG context, it has been answered in the negative. For example, the difference between interpretations (2-a) and (2-b) is not taken to correspond to any difference in the structure of (1) at any syntactic level of description. Instead, in the framework of Glue semantics, lexical and syntactic information from a sentence contribute a collection of linear logic premises, and every interpretation the sentence has corresponds to a different proof constructible from those premises.

In the case of (1), the premises contributed are shown in (3), both before and after instantiation based on its f-structure as shown in Figure 1.[1,2]

$$f : \begin{bmatrix} \textsc{pred} & \text{'guard'} \\ \textsc{subj} & g : \begin{bmatrix} \textsc{pred} & \text{'police officer'} \\ \textsc{spec} & i : \begin{bmatrix} \textsc{pred} & \text{'a'} \end{bmatrix} \end{bmatrix} \\ \textsc{obj} & h : \begin{bmatrix} \textsc{pred} & \text{'exit'} \\ \textsc{spec} & j : \begin{bmatrix} \textsc{pred} & \text{'every'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 1: F-structure of (1)

(3) $a \rightsquigarrow \lambda P.\lambda Q.\exists x.Px \wedge Qx : ((\textsc{spec}\uparrow) \multimap \uparrow) \multimap (((\textsc{spec}\uparrow) \multimap \%A) \multimap \%A)$

$\%A = ((\textsc{path}\uparrow)\ \textsc{spec}^*)$

$\Rightarrow \lambda P.\lambda Q.\exists x.Px \wedge Qx : (g \multimap i) \multimap ((g \multimap \%A) \multimap \%A)$

police officer $\rightsquigarrow \text{officer}' : \uparrow \multimap (\uparrow \textsc{spec})$

$\Rightarrow \text{officer}' : g \multimap i$

guards $\rightsquigarrow \text{guard}' : (\uparrow \textsc{subj}) \multimap ((\uparrow \textsc{obj}) \multimap \uparrow)$

$\Rightarrow \text{guard}' : g \multimap (h \multimap f)$

[1] I have chosen to define meaning constructors based on f-structure instead of s-structure, which is more common in the Glue literature. This choice has been made for simplicity of exposition and does not affect the argument at any point.

[2] Throughout this paper, expressions of the lambda calculus are given in $\beta\eta$-normal form. So for example, officer$' \equiv_\eta \lambda x.\text{officer}'x$ and guard$' \equiv_\eta \lambda x.\lambda y.\text{guard}'xy$.

$$
\begin{aligned}
&\textit{every} \rightsquigarrow \lambda P.\lambda Q.\forall y.Py \rightarrow Qy : ((\text{SPEC}\uparrow) \multimap \uparrow) \multimap \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad (((\text{SPEC}\uparrow) \multimap \%B) \multimap \%B) \\
&\qquad\qquad \%B = ((\text{PATH}\uparrow)\ \text{SPEC}^*) \\
&\qquad \Rightarrow \lambda P.\lambda Q.\forall y.Py \rightarrow Qy : (h \multimap j) \multimap ((h \multimap \%B) \multimap \%B) \\
&\textit{exit} \rightsquigarrow \text{exit}' : \uparrow \multimap (\uparrow \text{SPEC}) \\
&\quad \Rightarrow \text{exit}' : h \multimap j
\end{aligned}
$$

The proofs corresponding to surface scope and inverse scope interpretations are then shown in Figures 2 and 3 respectively.[3] Note that the lexical entries for the two determiners include local names ($\%A$ and $\%B$ respectively) specified by an inside-out functional uncertainty (IOFU) PATH.[4] For the time being we can take PATH to be maximally unconstrained, i.e. GF*. In any case, the only sensible value for either $\%A$ or $\%B$ in Figure 1 is f, and this is reflected in the two proofs.

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{\left[\begin{smallmatrix} x : \\ g \end{smallmatrix}\right]^1 \quad \begin{matrix}\text{guard}' : \\ g \multimap (h \multimap f)\end{matrix}}{\text{guard}'x : h \multimap f}
\quad
\dfrac{\begin{matrix}\text{exit}' : \\ h \multimap j\end{matrix} \quad \begin{matrix}\text{every}' : \\ (h \multimap j) \multimap \\ ((h \multimap f) \multimap f)\end{matrix}}{\text{every}'\text{exit}' : (h \multimap f) \multimap f}
}{\text{every}'\text{exit}'(\text{guard}'x) : f}
}{\lambda x.\text{every}'\text{exit}'(\text{guard}'x) : g \multimap f}\,1
\quad
\dfrac{\begin{matrix}\text{officer}' : \\ g \multimap i\end{matrix} \quad \begin{matrix}\text{a}' : \\ (g \multimap i) \multimap \\ ((g \multimap f) \multimap f)\end{matrix}}{\text{a}'\text{officer}' : (g \multimap f) \multimap f}
}{\begin{matrix}\text{a}'\text{officer}'(\lambda x.\text{every}'\text{exit}'(\text{guard}'x)) : f \\ \equiv \exists x.\text{officer}'x \wedge \forall y.\text{exit}'y \rightarrow \text{guard}'xy : f\end{matrix}}
$$

Figure 2: Proof for the surface scope interpretation of (1)

It is a major attraction of the Glue approach to semantic composition that it accounts for scope ambiguity like this, without the need for any additional assumptions such as are required by all the other theories already mentioned. However, that advantage can be seen as a disadvantage in cases where we do not see the same scope ambiguity as in (1).

---

[3]These proofs have been pared down in two ways to save space:

- The meaning representations of the two quantifiers have been written as a$'$ and every$'$, and only expanded in the final step.
- The inferential steps are not fully annotated. No annotation at all means that $\multimap$ elimination has been applied, and the number n means that $\multimap$ introduction has been applied, discharging the hypothesis numbered n.

[4] I follow the suggestion of Andrews (2010) and use IOFU, rather than quantification in the linear logic fragment, to specify scope level. The reason is mainly for cleanness of presentation, given that quantification will be introduced for other reasons later, although a potential advantage of this approach emerges in Section 1.3.

$$
\dfrac{
  \dfrac{
    \dfrac{
      \dfrac{
        \begin{bmatrix} y : \\ h \end{bmatrix}^2
        \quad
        \dfrac{
          \begin{bmatrix} x : \\ g \end{bmatrix}^1
          \quad
          \begin{array}{c} \text{guard}' : \\ g \multimap (h \multimap f) \end{array}
        }{
          \text{guard}'x : h \multimap f
        }
      }{
        \text{guard}'xy : f
      }
    }{
      \lambda x.\text{guard}'xy : g \multimap f
    }\,1
    \quad
    \dfrac{
      \begin{array}{c} \text{officer}' : \\ g \multimap i \end{array}
      \quad
      \begin{array}{c} \text{a}' : \\ (g \multimap i) \multimap \\ ((g \multimap f) \multimap f) \end{array}
    }{
      \text{a}'\text{officer}' : (g \multimap f) \multimap f
    }
  }{
    \dfrac{\text{a}'\text{officer}'(\lambda x.\text{guard}'xy) : f}{\lambda y.\text{a}'\text{officer}'(\lambda x.\text{guard}'xy) : h \multimap f}\,2
  }
  \quad
  \dfrac{
    \begin{array}{c} \text{exit}' : \\ h \multimap j \end{array}
    \quad
    \begin{array}{c} \text{every}' : \\ (h \multimap j) \multimap \\ ((h \multimap f) \multimap f) \end{array}
  }{
    \text{every}'\text{exit}' : (h \multimap f) \multimap f
  }
}{
  \begin{array}{l} \text{every}'\text{exit}'(\lambda y.\text{a}'\text{officer}'(\lambda x.\text{guard}'xy)) : f \\ \equiv \forall y.\text{exit}'y \rightarrow \exists x.\text{officer}'x \wedge \text{guard}'xy : f \end{array}
}
$$

Figure 3: Proof for the inverse scope interpretation of (1)

1.2 Scope rigidity

To take just one example, the German translation of (1) given in (4) does not show the same ambiguity as (1). Specifically, it **only** has the surface scope interpretation (2-a).

(4) Ein Polizist bewacht jeden Ausgang.
a.NOM police officer guards every.ACC exit

Ein	Polizist	bewacht	jeden	Ausgang.
a.NOM	police officer	guards	every.ACC	exit

This is a general property of German transitive clauses where the subject precedes the object, although not of clauses where the object scrambles over the subject. For example, (5) constrasts with (4) in not being scope-rigid. The argument structure is the same (in particular, exits are still being guarded) but both the surface scope (2-a) and the inverse scope (2-b) interpretation are possible again.

(5)

Jeden	Ausgang	bewacht	ein	Polizist.
every.ACC	exit	guards	a.NOM	police officer

For reference, partial c-to-f-structure mappings for (4) and (5) are shown in Figures 4 and 5 respectively.

In fact, we do not have to go outside English to find cases where quantifier scope is more rigid than Glue would lead us to expect. For example, quantifier scope is 'frozen' in double object constructions: (6) **only** has the interpretation represented in (6-a), not that represented in (6-b).

(6) Hilary gave a student every grade.

a. $\Rightarrow \exists y.\text{student}'y \wedge \forall x.\text{grade}'x \rightarrow \text{give}'\text{hilary}'xy$

b. $\not\Rightarrow \forall x.\text{grade}'x \rightarrow \exists y.\text{student}'y \wedge \text{give}'\text{hilary}'xy$

The purpose of this paper is to describe an extension to Glue semantics that will enable us to account both for the ambiguity of sentences like (1) and (5), and for the non-ambiguity of sentences like (4) and (6). But first, I need to set aside a potential red herring.

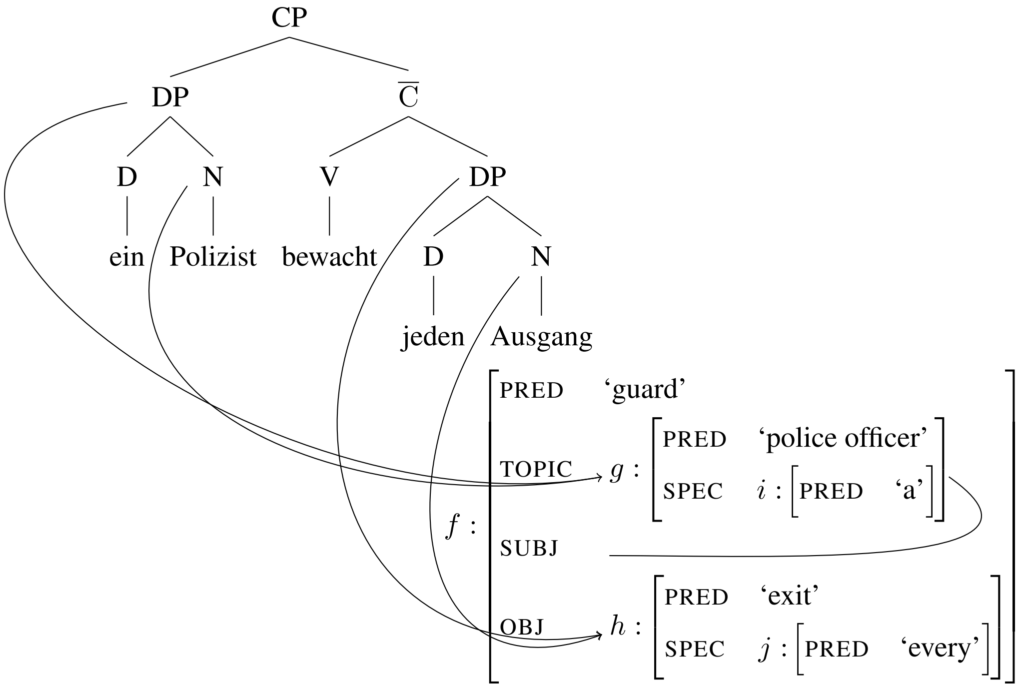


Figure 4: C- to f-structure of (4)

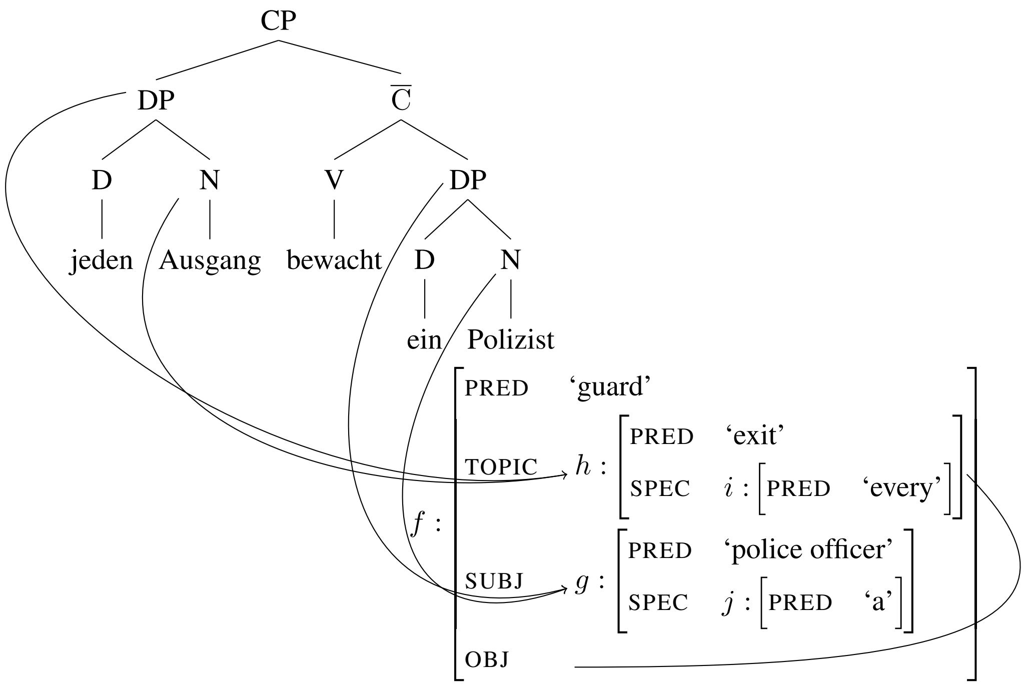


Figure 5: C- to f-structure of (5)

1.3 Not scope 'islands'

There are other examples of scope non-ambiguity that are not so hard to account for in Glue. For example, (7) only has the interpretation shown in (7-a), not that given in (7-b).

(7) If every student passes, the lecturer will be happy.

a. $\Rightarrow (\forall y.\text{student}'y \rightarrow \text{pass}'y) \rightarrow \text{happy}'(\imath x.\text{lecturer}'x)$

b. $\not\Rightarrow \forall y.\text{student}'y \rightarrow (\text{pass}'y \rightarrow \text{happy}'(\imath x.\text{lecturer}'x))$

That is to say, in the interpretation of (7), 'every' cannot take wider scope than 'if'.

$$f : \left[\begin{array}{ll} \textsc{pred} & \text{`happy'} \\ \textsc{subj} & [\text{``the lecturer''}] \\ \textsc{adj} & \left\{ g : \left[\begin{array}{ll} \textsc{pred} & \text{`pass'} \\ \textsc{compform} & \text{if} \\ \textsc{subj} & h : \left[\begin{array}{ll} \textsc{pred} & \text{`student'} \\ \textsc{spec} & i : \left[\begin{array}{ll} \textsc{pred} & \text{`every'} \end{array}\right] \end{array}\right] \end{array}\right] \right\} \end{array}\right]$$

Figure 6: F-structure of (7)

An explanation of this datum becomes obvious once we inspect the f-structure of (7) given in Figure 6 in combination with the lexical semantics of *every* as assumed in (3). To get the unavailable interpretation (7-b), the local name $\%B$ would have to resolve to f. This can be ruled out by making PATH suitably constrained—by not allowing it to pass through ADJ, for example.[5] $\%B$ would then have to resolve to g, which would allow for the derivation of (7-a) but not (7-b), as desired.

However, this kind of explanation is not available in the cases we're interested in. Take (4), the f-structure of which is shown in Figure 4, and assume that the meaning constructor for *jeden* is the same as for *every*. In Figure 4, as in Figure 1, the only way to get an interpretation of **either** the surface **or** (empirically unavailable) inverse scope interpretation is for $\%B$ to resolve to f. (4) and (6) are both examples of intra-clausal scope rigidity, and so in these cases the unwanted readings cannot be ruled out by constraining IOFU paths.

2 Previous work

I am aware of two existing proposals for constraining scope ambiguity in LFG+Glue: Crouch & van Genabith (1999) and Cook & Payne (2006).

[5] As hinted in Footnote 4, this explanation depends on the use of IOFU, rather than linear logic quantification, to fix scope level. If quantification is preferred for this purpose, an alternative explanation of scope islands could be to use extra modalities in the linear logic fragment, as in Gotham (2017).

2.1 Node orderings

According to Crouch & van Genabith (1999), in addition to meaning constructors, the linguistic system may also contribute certain admissibility conditions on linear logic proofs. Proofs failing to meet those conditions would then in some sense be filtered out. For example, the unavailable inverse scope interpretation of (4) could be ruled out by giving the main verb the (partial) lexical entry shown in (8).[6]

(8) *bewacht* V

$(\uparrow \text{PRED}) = \text{‘guard’}$
$\text{guard}' : (\uparrow \text{SUBJ}) \multimap ((\uparrow \text{OBJ}) \multimap \uparrow)$
$\boxed{(\uparrow \text{SUBJ}) <_f (\uparrow \text{OBJ}) \Rightarrow (\uparrow \text{SUBJ}) \succ (\uparrow \text{OBJ})}$

The consequent of the $\boxed{\text{boxed}}$ constraint in (8) is a **node ordering**. The intended interpretation of it is that in every licit linear logic derivation, the node labelled ($\uparrow$ SUBJ) must be ordered higher than that labelled ($\uparrow$ OBJ). Simplifying somewhat, for a linear logic formula x to be ordered higher than another y in a derivation d means that no instance of y occurs strictly lower down in d than every instance of x (Crouch & van Genabith, 1999, 132). So for example, if we apply (8) to Figure 4 or 5, we end up with

$$g <_f h \Rightarrow g \succ h$$

Since g f-precedes h in Figure 4, we then look at the candidate derivations in Figures 2 and 3.[7] We then see that the derivation in Figure 3 violates the constraint: there is an instance of h occurring strictly lower down than every instance of g. Therefore, this derivation is ruled out by the node ordering, and scope rigidity in enforced.

2.2 Objections to node orderings

2.2.1 Reifying representations?

As mentioned above, node orderings are constraints on derivations. But what, exactly, is a derivation? To inspect (8), in which the node ordering appears in a constraining equation, it seems that derivations have to be treated as levels in the projection architecture like c- or f-structure and that, just as with c- or f-structure, we can impose properly linguistic constraints on that structure:

> A derivation is a tree-like structure of sequents [. . .] Represent derivations $\mathcal{D}$ as triples $\langle S, >_S, \$\rangle$ where S is the set of points in the tree,

[6]I have had to slightly reconstruct the proposal of Crouch & van Genabith (1999) around other assumptions made in this paper. I do believe that this is a fair representation of their proposal.

[7]These are candidate derivations for (4) just as much as for (1) because the f-structure of (4) (Figure 4) does not differ from that of (1) (Figure 1) in any way that affects the contribution of meaning constructors.

> $>_S$ is a transitive , asymmetric ordering over them, and \$ is a function mapping the points onto their corresponding sequents. (Crouch & van Genabith, 1999, 131)

But from a logical and conceptual perspective this is deeply unsatisfactory, since it ties our use of the logic to a particular format for writing proofs out: namely, one with this kind of 'tree-like structure'. But, as Corbalán & Morrill (2016, fn. 4, emphasis mine) put it,

> Gentzen calculus, labelled and unlabelled natural deductions, proof nets, categorical calculus, etc. are all of repute, all have their respective advantages and disadvantages, and are **all notations for the same theory**.

Put differently, natural deduction derivations are representations of proofs, not the proofs themselves. The definition that Crouch & van Genabith (1999) provide requires us to write out proofs in a particular natural deduction format and not, for example, in sequent calculus or with proof nets.

Now, it should be pointed out that in all likelihood an equivalent notion of node ordering could be defined for all these other proof formats.[8] The real issue, though, is that if we have properly linguistic constraints on the form of derivations, we are not really working with proofs and therefore not doing logic any more. This clashes with an appealing (to me at least) picture according to which linear logic inference takes linguistic input and then operates entirely on its own terms, without the linguistic system needing to see the internal structure of any inferences.

2.2.2 Generate and filter?

Even if one sees no problem with treating derivations as linguistic objects in this way, one might well view node orderings with with some suspicion. The boxed clause in (8), for example, does not function like a normal constraining equation. There is no minimal linear logic derivation against which it has to be checked, on pain of failure—if there were, there would be no scope ambiguity! Instead, as mentioned above, the picture is that all logically-acceptable derivations are produced, but then (potentially) some are discarded.

What we need for a Glue-based theory of scope rigidity is a method for assigning linear logic formulae to lexical items such that all and only the desired interpretations of a sentence have a corresponding proof, rather that filtering out proofs by non-logical means based on derivations. Section 3 presents such a method.

[8]In fact, I have done just this in that the proof format used in this paper is not exactly that used by Crouch & van Genabith (1999).

2.3 Partition of meaning representations

In the theory of Cook & Payne (2006), topicality is not a grammatical function but rather a binary-valued f-structure attribute $\pm$T. Along with $\pm$N(ew) and $\pm$C(ontrastive), it constitutes the presence of information structure in f-structure. Word order in German is determined by the interaction of Optimality-Theoretic constraints referencing this information structure, as well as the GF hierarchy. The upshot of those constraints for a transitive clause is that, if the subject is $+$T and the object is $-$T then either of (4) or (5) is a possible word order, depending on other factors. But if the object is $+$T and the subject is $-$T, then only the scrambled order in (5) is possible. On the side of interpretation, there is a constraint to the effect that $+$T constituents must take scope over $-$T constituents; together, these constraints predict that (5) is scopally ambiguous but (4) is scope-rigid.

When we look at the implementation of that constraint on interpretation, we see that it is similar in spirit to Crouch & van Genabith's (1999) proposal. Instead of a constraint on the form of derivations, it is a constraint on the form of meaning representations. The constraint, called a partition, is shown in (9).

(9) $+\mathrm{T}(-\mathrm{T})$

This states that, in the final meaning representation of the clause, the meaning of any $-$T constituent must be contained in something that the meaning of a $+$T constituent is applied to. Given that the word order produced in (4) requires that a′officer′ be associated with $+$T and every′exit′ with $-$T, this predicts that (10-a) ($\equiv$ (2-a)) is a permitted meaning representation, but (10-b) ($\equiv$ (2-b)) is not.[9]

(10) a. $\text{a}'\text{officer}'(\lambda x.\text{every}'\text{exit}'(\text{guard}'x))$
b. $\text{every}'\text{exit}'(\lambda y.\text{a}'\text{officer}'(\lambda x.\text{guard}'xy))$

2.4 Objections

While Crouch & van Genabith's (1999) proposal reifies representations of proofs, Cook & Payne's (2006) proposal reifies representations of meanings. As Montague (1973) stressed, the step of translating natural language into a logical language like the lambda calculus, which in turn is interpreted in a model, should in principle be dispensable. The model-theoretic interpretations of (10-a) and (10-b) do not have any internal structure that could be used to distinguish them in the way envisaged in (9).

An alternative interpretation of Cook & Payne's (2006) proposal is as a constraint not on meaning representations but on the final step of how they are put together, given that application corresponds to $\multimap$ elimination by the Curry-Howard correspondence. Interpreted like that, it would amount to essentially the same thing as Crouch & van Genabith's (1999) proposal, as it would tie us to a proof format

[9] Cook & Payne (2006) in fact use quite different meaning representations, but that does not affect the substance of the argument.

that uses introduction and elimination rules (like natural deduction) rather than, say, left and right rules like sequent calculus.

3 A theory of scope rigidity

In general terms, the proposed account of scope rigidity proceeds as follow. First, we expand the fragment of linear logic used such that f-structure nodes are linear logic predicates (not formulae). We then use the arguments to those predicates to keep track of the order of application of quantifiers. Finally, in the lexicon, we set things up so that only by applying quantifiers in the desired order can a valid proof be constructed. This approach is inspired by work in Abstract Categorial Grammar (Pogodalla & Pompigne, 2012; Kanazawa, 2015).

3.1 Linear logic fragment

Given a set **P** of predicates (f-structure nodes) and a set **V** of variables (for which I wil use Greek letters), the fragment of linear logic used is as defined in (11).

(11)
$$\begin{array}{rcll} n & ::= & \mathbf{V} \mid 0 \mid \mathsf{s}n & \text{(terms)} \\ \phi, \psi & ::= & \mathbf{P}n \mid \phi \multimap \psi \mid \forall \mathbf{V}.\phi & \text{(formulae)} \end{array}$$

Formally, for much of what we want this fragment to do s can be an unanalyzed function symbol. However, for readability, in what follows we will treat it as the successor function and represent $\mathsf{s}0$ as 1, $\mathsf{s}(\mathsf{s}0)$ as 2 etc.

3.2 Lexicon

We can now enforce scope rigidity for (4) by means of the mini German lexicon shown in (12), assuming the f-structure in Figure 4. In order to save space and improve readability, I have written arguments to linear logic predicates and functions in subscript, e.g. instead of writing $g0$ I write g_0.

(12)
$$\begin{array}{l} \textit{ein} \rightsquigarrow \lambda P.\lambda Q.\exists x.Px \wedge Qx : \forall \iota.((\textsc{spec}\uparrow)_0 \multimap \uparrow_0) \multimap \\ \qquad\qquad\qquad\qquad (((\textsc{spec}\uparrow)_{\mathsf{s}\iota} \multimap \%A_{\mathsf{s}\iota}) \multimap \%A_\iota) \\ \qquad \%A = ((\textsc{path}\uparrow)\ \textsc{spec}^*) \\ \qquad \Rightarrow \lambda P.\lambda Q.\exists x.Px \wedge Qx : \forall \iota.(g_0 \multimap i_0) \multimap \\ \qquad\qquad\qquad\qquad ((g_{\mathsf{s}\iota} \multimap \%A_{\mathsf{s}\iota}) \multimap \%A_\iota) \\ \textit{Polizist} \rightsquigarrow \text{officer}' : \uparrow_0 \multimap (\uparrow \textsc{spec})_0 \\ \qquad \Rightarrow \text{officer}' : g_0 \multimap i_0 \\ \textit{bewacht} \rightsquigarrow \text{guard}' : \forall \iota.\forall \eta.(\uparrow \textsc{subj})_\iota \multimap ((\uparrow \textsc{obj})_\eta \multimap \uparrow_\eta) \\ \qquad \Rightarrow \text{guard}' : \forall \iota.\forall \eta.g_\iota \multimap (h_\eta \multimap f_\eta) \end{array}$$

$$
\begin{aligned}
\textit{jeden} \rightsquigarrow{} & \lambda P.\lambda Q.\exists x.Px \wedge Qx : \forall\iota.((\textsc{spec}\uparrow)_0 \multimap {\uparrow}_0) \multimap \\
& \qquad\qquad\qquad (((\textsc{spec}\uparrow)_{\mathsf{s}\iota} \multimap \%B_{\mathsf{s}\iota}) \multimap \%B_\iota) \\
& \%B = ((\textsc{path}\uparrow)\ \textsc{spec}*) \\
\Rightarrow{} & \lambda P.\lambda Q.\forall y.Py \rightarrow Qy : \forall\iota.(h_0 \multimap j_0) \multimap \\
& \qquad\qquad\qquad ((h_{\mathsf{s}\iota} \multimap \%B_{\mathsf{s}\iota}) \multimap \%B_\iota) \\
\textit{Ausgang} \rightsquigarrow{} & \mathrm{exit}' : {\uparrow}_0 \multimap (\uparrow \textsc{spec})_0 \\
\Rightarrow{} & \mathrm{exit}' : h_0 \multimap j_0
\end{aligned}
$$

3.3 Derivations

With the lexicon shown in (12), the surface scope interpretation can be derived, as shown in Figure 7, but the inverse scope interpretation cannot. Figure 8 shows one failed attempt to do so. Intuitively, the effect of the lexicon is to introduce a 'counter' as the argument to linear logic predicates, which forces quantifiers to apply in a particular order.

$$
\dfrac{
  \dfrac{
    \dfrac{
      [x : g_1]^1 \quad
      \dfrac{\mathrm{guard}' : \forall\iota.\forall\eta.g_\iota \multimap (h_\eta \multimap f_\eta)}{\mathrm{guard}' : g_1 \multimap (h_2 \multimap f_2)}\,\forall_E
    }{\mathrm{guard}'x : h_2 \multimap f_2}
    \quad
    \begin{array}{c}\textit{jeden Ausgang}\\ \vdots \\ \mathrm{every}'\mathrm{exit}' : \\ (h_2 \multimap f_2) \multimap f_1\end{array}
  }{
    \dfrac{\mathrm{every}'\mathrm{exit}'(\mathrm{guard}'x) : f_1}{\lambda x.(\mathrm{every}'\mathrm{exit}'(\mathrm{guard}'x)) : g_1 \multimap f_1}\,1
  }
  \quad
  \begin{array}{c}\textit{Ein Polizist}\\ \vdots \\ \mathrm{a}'\mathrm{officer}' : \\ (g_1 \multimap f_1) \multimap f_0\end{array}
}{\mathrm{a}'\mathrm{officer}'(\lambda x.(\mathrm{every}'\mathrm{exit}'(\mathrm{guard}'x))) : f_0}
$$

Figure 7: Proof for the (surface scope) interpretation of (4)

$$
\dfrac{
  \dfrac{
    \dfrac{
      [y : h_1]^2 \quad
      \dfrac{
        [x : g_2]^1 \quad
        \dfrac{\mathrm{guard}' : \forall\iota.\forall\eta.g_\iota \multimap (h_\eta \multimap f_\eta)}{\mathrm{guard}' : g_2 \multimap (h_1 \multimap f_1)}\,\forall_E
      }{\mathrm{guard}'x : h_1 \multimap f_1}
    }{\mathrm{guard}'xy : f_1}
  }{\lambda x.\mathrm{guard}'xy : g_2 \multimap f_1}\,1
  \quad
  \begin{array}{c}\textit{ein Polizist}\\ \vdots \\ \mathrm{a}'\mathrm{officer}' : \\ (g_2 \multimap f_2) \multimap f_1\end{array}
}{}\,*
$$

Figure 8: Failed attempt to derive an inverse scope interpretation for (4)

The lexical entries for the determiners mean that applying a quantifier reduces the counter by one (from $\mathsf{s}\iota$ to ι). It follows that for quantifier Q_1 to immediately outscope quantifier Q_2 (apply immediately after it), you have to set the counter for Q_1 to one lower than for Q_2. Therefore, to get the inverse scope reading of (4), you would have to set the counter for the subject position one higher than for the object position; this is how the failed attempt in Figure 8 starts. However, the lexical entry for the verb guarantees that if you do that, no proof can be constructed: it sets the counter for the clause to the same as for the object position, which makes it impossible to apply the subject quantifier first.

4 Loosening the restrictions

In the above example, scope rigidity is enforced by a combination of the lexical entries for the determiners and the verb. The restriction could, therefore, be relaxed by tweaking either lexical entry. For instance, assuming that English uses the same fragment of linear logic for its syntax-semantics interface as German (which seems reasonable), the English data can be accounted for by assigning to *guards* the meaning constructor shown in (13), and otherwise importing the translation of the German lexicon.

(13) $\text{guard}' : \forall\iota.\forall\eta.\forall\kappa.(\uparrow \textsc{subj})_\iota \multimap ((\uparrow \textsc{obj})_\eta \multimap {\uparrow}_\kappa)$

The meaning constructor assigned to *guards* in (13) differs from that assigned to *bewacht* in (12) in that in (13) the counter for the clause is not tied to the same value as either argument position. It could be instantiated with the same value as the object, enabling a surface scope interpretation, or to the subject, enabling an inverse scope interpretation.

However, in many instances we want the flexibility we allow for to be more fine-grained than this.

4.1 German scrambling

So far, the theory in Section 3 does not distinguish between (4) and (5); they are predicted to both only have the interpretation (2-a). What it seems that we would like to have is conditional introduction of meaning constructors: some statement to the effect that *bewacht* introduces the meaning constructor given in (12) if its subject precedes its object, and the English-style meaning constructor (13) otherwise. I do not see a way to do that directly in the formal architecture of LFG+Glue, but there is a workaround: give a German transitive verb the lexical entry shown in (14),

(14) *bewacht* V

$(\uparrow \textsc{pred}) = \text{`guard'}$
$\text{guard}' : \forall\iota.\forall\eta.(\uparrow \textsc{subj})_\iota \multimap ((\uparrow \textsc{obj})_\eta \multimap {\uparrow}_\eta)$
$(@\textsc{reset})$

where RESET is the template defined in (15).

(15) RESET := $(\uparrow \text{OBJ}) <_f (\uparrow \text{SUBJ})$

$$\lambda p.p : \forall \iota.\forall \eta.\ \uparrow_\iota \multimap \uparrow_\eta$$

The RESET template is so called because it can be used to reset the counter. As specified in (14), the introduction of this template is optional. Clearly, if the subject f-precedes the object, as in Figure 4, then calling RESET would cause failure. So in the case of (4) RESET cannot be called, and scope rigidity is enforced as described in Section 3.

However, if the object f-precedes the subject, as in Figure (5), then RESET may or may not be called. If it is called, then both scope orderings are possible since the counter can be changed. Figure 9 shows a derivation of the (2-b) interpretation of (5) based on Figure 5, picking up from a decision point in the failed attempt for (4) shown in Figure 8.

$$
\dfrac{
  \dfrac{
    \dfrac{
      \dfrac{
        \dfrac{
          \dfrac{\overset{\text{(see figure 8)}}{\begin{bmatrix} y : \\ h_1 \end{bmatrix}^2, \begin{bmatrix} x : \\ g_2 \end{bmatrix}^1, \textit{bewacht}} \\ \vdots}{\text{guard}'xy : f_1}
          \quad
          \dfrac{\overset{(\text{RESET})}{\lambda p.p : \forall \iota.\forall \eta.f_\iota \multimap f_\eta}}{\lambda p.p : f_1 \multimap f_2}\ \forall_E
        }{\dfrac{\text{guard}'xy : f_2}{\lambda x.\text{guard}'xy : g_2 \multimap f_2}\ 1}
        \quad
        \dfrac{\textit{ein Polizist} \\ \vdots}{\text{a}'\text{officer}' : (g_2 \multimap f_2) \multimap f_1}
      }{\text{a}'\text{officer}'(\lambda x.\text{guard}'xy) : f_1}
    }{\lambda y.\text{a}'\text{officer}'(\lambda x.\text{guard}'xy) : h_1 \multimap f_1}\ 2
    \quad
    \dfrac{\textit{jeden Ausgang} \\ \vdots}{\text{every}'\text{exit}' : (h_1 \multimap f_1) \multimap f_0}
  }{}
}{
\begin{array}{c}
\text{every}'\text{exit}'(\lambda y.\text{a}'\text{officer}'(\lambda x.\text{guard}'xy)) : f_0 \\
\equiv \forall y.\text{exit}'y \rightarrow \exists x.\text{officer}'x \wedge \text{guard}'xy : f_0
\end{array}
}
$$

Figure 9: An interpretation of (5) that is not available for (4)

4.2 English double object constructions

In order to analyze scope freezing in English double object constructions, we have to generalize from the example in (6) by seeing what interpretations are available when there is a quantifier in subject position too, as in (16).

(16) A teacher gave most students every grade.

Judgements get a little unclear for examples like this, but Bruening (2001) reports the results shown in (17), and I will proceed on that basis. In summary, the secondary object may not take wider scope than the primary object, but otherwise the relative scope of the three quantifiers is free.[10]

(17) a. (16) $\Rightarrow \exists x.\text{teacher}'x \wedge \text{most}'\text{student}'(\lambda z.\forall y.\text{grade}'y \rightarrow \text{give}'xyz)$

[10]Just to be clear: give$'xyz$ should be interpreted as saying that x gives y to z (or equivalently, the x gives z y).

b. (16) $\Rightarrow$ most′student′$(\lambda z.\exists x.$teacher′$x \wedge \forall y.$grade′$y \rightarrow$ give′$xyz)$
c. (16) $\Rightarrow$ most′student′$(\lambda z.\forall y.$grade′$y \rightarrow \exists x.$teacher′$x \wedge$ give′$xyz)$
d. (16) $\not\Rightarrow$ $\exists x.$teacher′$x \wedge \forall y.$grade′$y \rightarrow$ most′student′$(\lambda z.$give′$xyz)$
e. (16) $\not\Rightarrow$ $\forall y.$grade′$y \rightarrow \exists x.$teacher′$x \wedge$ most′student′$(\lambda z.$give′$xyz)$
f. (16) $\not\Rightarrow$ $\forall y.$grade′$y \rightarrow$ most′student′$(\lambda z.\exists x.$teacher′$x \wedge$ give′$xyz)$

An attempt to account for the data in (17), which undergenerates slightly, is to assign to *gave* the meaning constructor shown in (18).

(18) give′ : $\forall\iota.\forall\eta.\forall\kappa.(\uparrow \text{SUBJ})_\iota \multimap ((\uparrow \text{OBJ}_\theta)_\eta \multimap ((\uparrow \text{OBJ})_\kappa \multimap \uparrow_\eta))$

Using (18) forces the secondary object to take narrowest scope, and otherwise leaves the relative scope of the quantifiers free. That correctly rules out the unavailable readings of (16), and predicts two of the three available ones, but incorrectly rules out the interpretation shown in (17-c).

There are, of course, various ways of tackling this shortcoming.[11] The one I will present here involves once more expanding the fragment of linear logic used, by adding to the language defined in Section 3.1 a ternary function symbol c, to be interpreted as shown in (19).[12]

(19) $$\mathsf{c}mno = \begin{cases} m \text{ if } m > n > o \\ n \text{ otherwise} \end{cases}$$

We can now capture the data in (17) by assigning *gave* the meaning constructor shown in (20).

(20) give′ : $\forall\iota.\forall\eta.\forall\kappa.(\uparrow \text{SUBJ})_\iota \multimap ((\uparrow \text{OBJ}_\theta)_\eta \multimap ((\uparrow \text{OBJ})_\kappa \multimap \uparrow_{\mathsf{c}\iota\eta\kappa}))$

For illustration, derivations of the interpretations (17-b) and (17-c) are shown in Figures 11 and 12, respectively, based on the f-structure in Figure 10. By contrast, a failed attempt to derive (17-d) is shown in Figure 13.

$$f : \begin{bmatrix} \text{PRED} & \text{‘give’} \\ \text{SUBJ} & g : \begin{bmatrix}\text{“a teacher”}\end{bmatrix} \\ \text{OBJ} & h : \begin{bmatrix}\text{“most students”}\end{bmatrix} \\ \text{OBJ}_\theta & i : \begin{bmatrix}\text{“every grade”}\end{bmatrix} \end{bmatrix}$$

Figure 10: F-structure of (16)

[11] Some will be hinted at in Section 5.

[12] We now have to take seriously that s is the successor function, since we have $>$ at the meta-level.

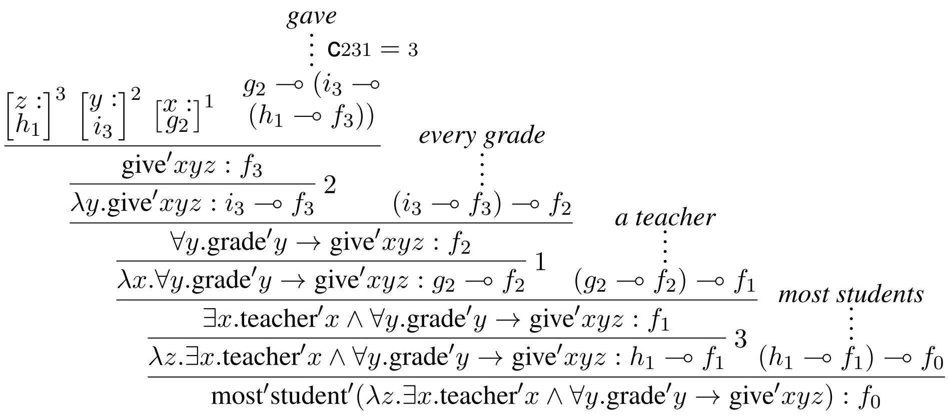


Figure 11: Proof of (17-b)

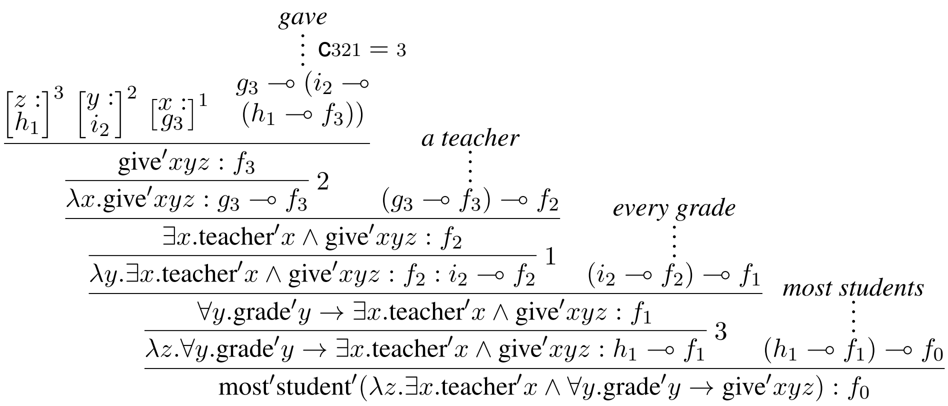


Figure 12: Proof of (17-c)

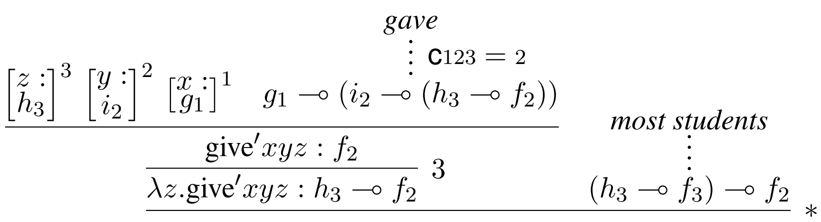


Figure 13: A failed attempt to derive (17-d)

5 Discussion

This has primarily been a theoretical paper, describing and applying certain formal tools for constraining scope ambiguity in LFG+Glue. I have argued that those tools are conceptually preferable to the ones previously put forward in the literature. In a nutshell, the account of scope rigidity presented in this paper is that a verb may specify, in its lexical entry, which of its arguments must or may take narrowest scope. Ideally, at least for the cases considered in this paper, we would like the specifications to be more general than that, such that they would apply to every transitive (for German) or ditransitive (for English) verb in the language. Presumably the desired effect can be achieved, at least at the descriptive level, with the use of templates (Dalrymple et al., 2004).

Of course, this paper has not even scratched the surface of the empirical data to be accounted for in a theory of scope rigidity. One major limitation is that we have only considered cases where the constraints on scope ordering can be described purely in terms of grammatical functions. But there are cases where the possibility of scope ambiguity seems to depend on precisely which quantifiers occupy which of a verb's argument positions. For example, while English generally allows inverse scope in simple transitive sentences, it is not clear that this is always possible when there is a monotone-decreasing quantifier in object position, as in (21).

(21) Most students completed fewer than three assignments.

The right approach to this datum (if it is one) may be to treat English more like German, only with a less constrained RESET-like template for loosening restrictions. Or alternatively, it may be more profitable to treat the quantifiers themselves as contributing the necessary restrictions, as in Fry's (1999) account of NPI licensing. Many details would need to be worked out, though: as Fry (1999) acknowledges, his theory has the power to force certain items (NPIs) to take scope under other items (licensors), but no power to require licensors to be more prominent than licensees at any level of grammatical description.

There are other possible alternatives that could be explored. For example, an account of the kind given for scope islands in Section 1.3 *could* be extended to scope rigidity, but it would require some quite drastic foundational changes to the LFG+Glue architecture: either by providing f-/s-structure with more internal structure (cf. Andrews (2018) on the relative scope of adjectives), or by having linear logic formulae read off c-structure instead. I cannot seen either of these options being popular.

References

Andrews, Avery D. 2010. Propositional glue and the projection architecture of LFG. *Linguistics and Philosophy* 33. 141–170. doi:10.1007/s10988-010-9079-9.

Andrews, Avery D. 2018. Sets, heads, and spreading in LFG. *Journal of Language Modelling* 6(1). 131–174. doi:10.15398/jlm.v6i1.175.

Barker, Chris & Pauline Jacobson (eds.). 2007. *Direct compositionality* (Oxford Studies in Theoretical Linguistics 14). Oxford: Oxford University Press.

Barker, Chris & Chung-chieh Shan. 2014. *Continuations and natural language* (Oxford Studies in Theoretical Linguistics 53). Oxford: Oxford University Press.

Bruening, Benjamin. 2001. QR obeys superiority: Frozen scope and ACD. *Linguistic Inquiry* 32(2). 233–273. doi:10.1162/00243890152001762.

Cook, Philippa & John Payne. 2006. Information structure and scope in German. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'06 conference, University of Konstanz*, Stanford, CA: CSLI Publications.

Cooper, Robin. 1983. *Quantification and syntactic theory* (Studies in Linguistics and Philosophy 21). Dordrecht: D. Reidel.

Corbalán, María Inés & Glyn Morrill. 2016. Overtly anaphoric control in type logical grammar. In Annie Foret, Glyn Morrill, Reinhard Muskens, Rainer Osswald & Sylvain Pogodalla (eds.), *Formal grammar* (Lecture Notes in Computer Science 9804), 183–199. Berlin, Heidelberg: Springer. doi:10.1007/978-3-662-53042-9_11.

Crouch, Richard & Josef van Genabith. 1999. Context change, underspecification and the structure of Glue language derivations. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar*, 117–189. Cambridge, MA: MIT Press.

Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'04 conference, University of Canterbury*, 199–208. Stanford, CA: CSLI Publications.

Fry, John. 1999. Proof nets and negative polarity licensing. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar*, 91–116. Cambridge, MA: MIT Press.

Gotham, Matthew. 2017. Glue semantics and locality. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'17 conference, University of Konstanz*, 230–242. Stanford, CA: CSLI Publications.

Jacobson, Pauline. 2002. The (dis)organization of the grammar. *Linguistics and Philosophy* 25. 601–626. doi:10.1023/A:1020851413268.

Jacobson, Pauline. 2014. *Compositional semantics*. Oxford: Oxford University Press.

Kanazawa, Makoto. 2015. Syntactic features for regular constraints and an approximation of directional slashes in abstract categorial grammars. In Yusuke Kubota & Robert Levine (eds.), *Empirical advances in categorial grammar*, 34–70. https://web.archive.org/web/20170705131723/https://www.u.tsukuba.ac.jp/~kubota.yusuke.fn/cg2015-proceedings.pdf. Last accessed 2019-10-07.

May, Robert. 1985. *Logical form* (Linguistic Inquiry Monographs 12). Cambridge, MA: MIT Press.

Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In Patrick Suppes, Julius Moravcsik & Jaakko Hintikka (eds.), *Approaches to natural language*, 221–242. Dordrecht: D. Reidel.

Pogodalla, Sylvain & Florent Pompigne. 2012. Controlling extraction in abstract categorial grammars. In Philippe de Groote & Mark-Jan Nederhof (eds.), *Formal grammar* (Lecture Notes in Computer Science 7395), 162–177. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-32024-8_11.

Tractability and Discontinuity

Ronald M. Kaplan
Stanford University

Jürgen Wedekind
University of Copenhagen

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 130–148

http://csli-publications.stanford.edu/LFG/2019

Keywords: tractability, undecidability, degree of discontinuity, functional domain, finite boundedness, nonconstructivity, LCFRS, linear context-free rewriting system, finite copying grammar, X-bar theory

Kaplan, Ronald M., & Wedekind, Jürgen. 2019. Tractability and Discontinuity. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 130–148. Stanford, CA: CSLI Publications.

Abstract

LFG rule systems that embody the explanatory principles of X-bar theory appear to allow for arbitrary repetitions of nodes, particularly complements and coheads, that map to the same f-structure. Such X-bar compliant grammars thus fail to meet the requirements for tractable computation that have been identified in recent work (Wedekind and Kaplan, to appear). This raises the question whether those grammars also fail of descriptive accuracy in that they provide for derivations with syntactic dependencies that are not attested in natural languages. We address what may be regarded as a discrepancy between explanatory and descriptive adequacy by imposing finite bounds on the degree of discontinuity of grammatical functions and the number of nodes in functional domains. We introduce additional filtering conditions on derivations, along the lines of Completeness and Coherence and the original nonbranching dominance prohibition, that make it possible to decide that derivations respect those bounds. Grammars that embody the principles of X-bar theory and satisfy these bounds and additional requirements are amenable to tractable processing.

1 Introduction

It is well known that the recognition/parsing problem is decidable for the LFG formalism restricted only by the nonbranching dominance condition (Kaplan and Bresnan, 1982), and the generation problem is decidable for arbitrary LFG grammars as long as the input f-structure is acyclic (Wedekind and Kaplan, 2012; Wedekind, 2014). But even with these restrictions these problems are also known to be computationally intractable in the worst case, as they still belong to the class of NP-complete problems (e.g. Berwick, 1982).

However, grammars for actual languages seem not to exploit all the mathematical power that the formalism makes available, as witnessed by the fact that parsing and generation systems, for example, the XLE system, have been constructed that are practical for broad-coverage grammars and naturally occurring sentences (Maxwell and Kaplan, 1996; Crouch et al., 2008). The implementations of these systems must be implicitly taking advantage of certain patterns of dependencies that are characteristic of linguistic grammars even if those properties have not yet been clearly articulated, and their computational consequences are not yet clearly understood and have not been explicitly coded.

Wedekind and Kaplan (to appear) (henceforth WK) define a subclass of LFG grammars with formal properties that, they suggest, are compatible with linguistic description but make possible the translation into equivalent grammars in the framework of linear context-free rewriting systems (LCFRS), a mildly context-sensitive grammatical formalism (Seki et al., 1991;

Kallmeyer, 2013). The recognition and emptiness problems of LCFRS grammars are decidable, and recognition is tractable in the sense that it requires time that is proportional to a polynomial (vs. an exponential) function of the length of the input. WK show that it is decidable whether an arbitrary LFG grammar in the original Kaplan and Bresnan (1982) formalism meets the requirements of the restricted subclass and thus admits of an LCFRS equivalent.

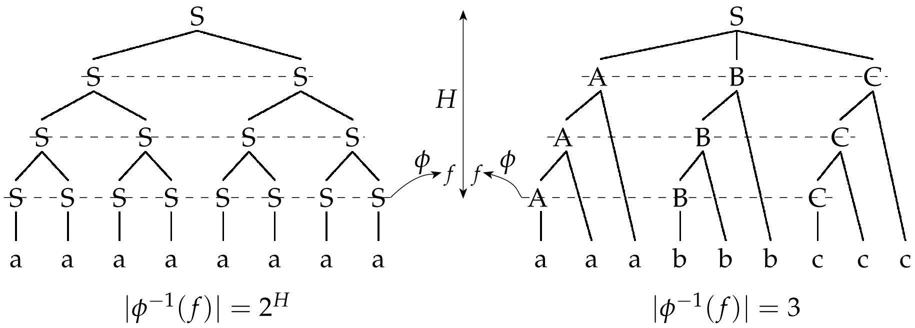
In this paper we summarize the formal properties that WK have identified, and we indicate informally why computational performance is sensitive to these particular restrictions. These restrictions include imposing finite upper bounds on the size of functional domains and the c-structure discontinuity of functional units. We observe, however, that these bounds are not realized by grammars whose rules conform to the schematic prescriptions of X-bar theory but are not otherwise constrained. We provide an LFG formalization that includes two extragrammatical parameters, the *degree of discontinuity* (Chomsky, 1953) and the *height of the functional domains*, and show that conditions for tractability are achieved if we allow only derivations that respect finite bounds on these parameters even though a language may not limit the number of modifers. Thus, if bounded derivations can account for all and only the sentences of natural languages, grammars can instantiate the explanatory principles of X-bar theory and still admit of tractable computation.

2 Tractable LFG grammars

WK take as a point of departure the severely restricted subclass of LFG grammars that were shown to be tractable by Seki et al. (1993). The functional annotations of these *finite-copying grammars* are severely restricted: they allow categories to be annotated only with at most one function assignment of the form $(\uparrow \text{ G}) = \downarrow$ and feature assignments of the form $(\uparrow \text{ A}) = v$ that assert feature values with a single attribute just on the mother nodes of c-structure expansions.

This notation is clearly unsuitable for linguistic description. It disallows the trivial $\uparrow = \downarrow$ annotations that mark the heads and coheads in the functional domain of a predicate, the $(\uparrow \text{ XCOMP SUBJ}) = (\uparrow \text{ OBJ})$ equations of functional control, and all other ways of relating the f-structures of different nodes. It also disallows longer attribute paths in the specification of mother-node feature values (e.g. $(\uparrow \text{ SUBJ NUM}) = \text{SG}$) and any direct specification of feature values on daughter nodes, as in $(\downarrow \text{ CASE}) = \text{ NOM}$.

In terms of formal expressiveness, on the other hand, the function-assignment annotations in these grammars do allow separate nodes of the c-structure to map to the same unit of f-structure and thus to share (and require consistency of) atom-value information in a very constrained way.



a. C-structure for grammar (1) b. C-structure for grammar (2)

Figure 1: Zipper nodes in depth-balanced c-structures

This specific type of structure sharing is occasionally referred to as "zipper" unification. That is, two distinct nodes n and n' can map to the same f-structure in a valid derivation (formally, $\phi(n) = \phi(n')$) if there is a node $\hat{n}$ dominating both of these nodes and the sequences of function-assignment annotations on the c-structure paths from $\hat{n}$ to n and n', respectively, are identical, that is, form a "zipper". As a significant consequence, structure-sharing nodes for finite-copying grammars must have the same c-structure depth. This is illustrated by the simple grammars in (1) and (2) and by the depth-balanced c-structures in Figure 1, where the dashed lines link collections of nodes that map through ϕ to the same f-structure.

(1) $\mathrm{S} \;\rightarrow\; \begin{array}{c}\mathrm{S}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \begin{array}{c}\mathrm{S}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \qquad\qquad \mathrm{S} \;\rightarrow\; \begin{array}{c}\mathrm{a}\\ (\uparrow \textsc{l}) = \#\end{array}$

(2) $\mathrm{S} \;\rightarrow\; \begin{array}{c}\mathrm{A}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \begin{array}{c}\mathrm{B}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \begin{array}{c}\mathrm{C}\\ (\uparrow \textsc{l}){=}\downarrow\end{array}$

$\mathrm{A} \;\rightarrow\; \begin{array}{c}\mathrm{A}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \mathrm{a} \qquad\qquad \mathrm{A} \;\rightarrow\; \begin{array}{c}\mathrm{a}\\ (\uparrow \textsc{l}) = \#\end{array}$

$\mathrm{B} \;\rightarrow\; \begin{array}{c}\mathrm{B}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \mathrm{b} \qquad\qquad \mathrm{B} \;\rightarrow\; \begin{array}{c}\mathrm{b}\\ (\uparrow \textsc{l}) = \#\end{array}$

$\mathrm{C} \;\rightarrow\; \begin{array}{c}\mathrm{C}\\ (\uparrow \textsc{l}){=}\downarrow\end{array} \quad \mathrm{c} \qquad\qquad \mathrm{C} \;\rightarrow\; \begin{array}{c}\mathrm{c}\\ (\uparrow \textsc{l}) = \#\end{array}$

While these grammars both meet the very stringent finite-copying notational restrictions, the difference in these structure-sharing configurations corresponds to a difference in computational complexity. For all derivations of grammar (1) the number of nodes that map to a given f-structure f ($= |\phi^{-1}(f)|$) is an exponential in the height H of those nodes, as illustrated in Figure 1a. In contrast, for all derivations of grammar (2) the number of nodes in a structure-sharing set is bounded by a constant (3 in this case) that is independent of their height (Figure 1b). Grammar (2) but not (1) meets the finite-boundedness property (3) that Seki et al. (1993) also included in the definition of finite-copying grammars.

(3) A grammar is **finitely bounded** if and only if there is a grammar-dependent constant d such that no more than d nodes map to the same f-structure element f in any derivation. That is, $|\phi^{-1}(f)| \leq d$.

They established that this is a decidable property of grammars with only function-assignment and mother-feature annotations and thus whether any particular grammar belongs to the finite-copying subclass. They further demonstrated that for any such finitely-bounded and notationally restricted grammar G there is an equivalent mildly context-sensitive grammar G' (an LCFRS). It follows that for this grammar class that the emptiness problem is decidable, the recognition problem is tractable, and parsing with G can be accomplished with the equivalent LCFRS G'. The key insight is that atom-based satisfiability can be tested in zipper domains of no more than d nodes, even when these are not adjacent in the c-structure. We regard the parameter d as formalizing the linguistic notion of the degree of discontinuity (Chomsky, 1953).

Finite boundedness is a necessary property for LCFRS equivalence and tractability. It is a sufficient property for finite-copying grammars, but further restrictions are required for grammars which include annotations in more elaborate formats. WK carefully extend the notation to allow for annotations that are commonly used in LFG accounts of syntactic phenomena and are thus more suitable for linguistic description. These extensions are introduced with formal restrictions that conserve the LCFRS equivalence and thus preserve the key advantages of that minimally expressive formalism. They first extend the notation to admit more elaborate atomic-value annotations on both mother and daughter nodes. This includes annotations such as $(\uparrow \text{SUBJ NUM}) = \text{SG}$ and $(\downarrow \text{NUM}) = \text{PL}$ and in general any annotations of the forms $(\uparrow \text{A B C} \ldots) = v$ and $(\downarrow \text{A B C} \ldots) = v$.

They also introduce a broader but still limited range of annotations that establish relationships between the f-structures associated with different nodes in a derivation. They allow trivial annotations $\uparrow = \downarrow$, and reentrancies of the following forms:[1]

[1]WK also allow set-membership annotations $(\downarrow \in (\uparrow \text{ADJ}))$ that are typically used for modifiers. Membership annotations only map single nodes to f-structures and do not propagate atomic feature values, and thus are inert with respect to zipper formation, satisfiability, and computational complexity. We return to this point below.

(4) $(\uparrow \text{F G}) = (\uparrow \text{H})$ functional control
$(\uparrow \text{F}) = (\uparrow \text{H})$ local-topic link
$(\downarrow \text{G}) = (\uparrow \text{H})$ SUBJ in XADJ control
$(\downarrow \text{G}) = (\downarrow \text{H})$ daughter sharing
$(\downarrow \text{G}) = \uparrow$ promotion
$(\uparrow \text{F}) = \uparrow$ mother cycle
$(\downarrow \text{G}) = \downarrow$ daughter cycle

These annotations allow designators with at most two grammatical functions, in keeping with the Principle of Functional Locality (Kaplan and Bresnan, 1982). WK hypothesize that this inventory of annotation-types is sufficient for natural language description, since the notation can account for function assignments, head and cohead identification, functional control, and agreement. This claim receives empirical support from an examination of the large-scale, broad-coverage Pargram grammars of English and German. There are only a few outliers in the 281 rules and 23,809 lexical entries of the English grammar, for example, and these appear only in still controversial accounts of oblique prepositional phrases and copular complements.

WK also demonstrate that this family of linguistically suitable annotations is also expressive enough to make the core computational problems—recognition (without the nonbranching dominance restriction), emptiness, and generation from cyclic inputs—undecidable.[2]

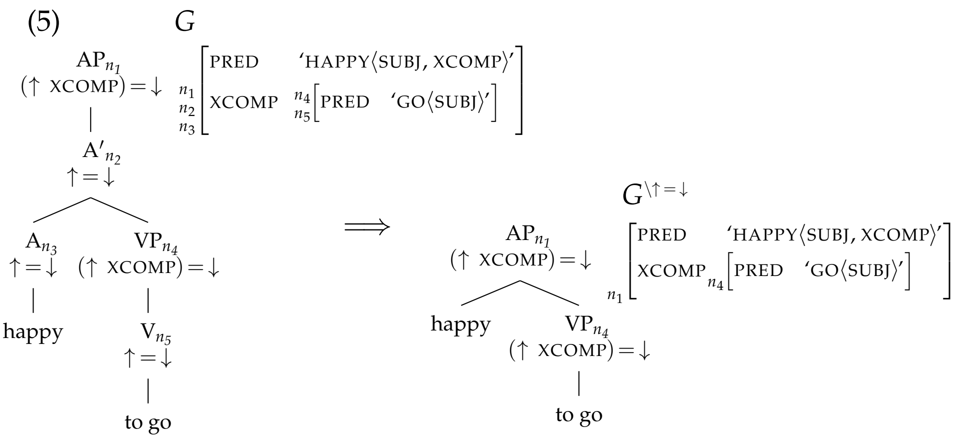
WK then consider a set of restrictions that together ensure LCFRS equivalence. They conjecture that these restrictions will not undermine the descriptive utility of the reentrancies and trivial annotations. WK first require, similar to Seki et al., that nonterminal categories are annotated with at most one function assignment of the form $(\uparrow \text{G}) = \downarrow$, and, furthermore, that trivial annotations and function assignments always appear in complementary distribution to keep separate the properties of a head and its complements.

Since finite boundedness is a necessary condition for LCFRS equivalence, clearly there must be a finite bound on the number of nodes in a functional domain. These nodes are annotated with $\uparrow = \downarrow$ annotations and typically carry information about heads, coheads, and the local morphosyntactic features of a functional unit. These $\uparrow = \downarrow$ annotations map all the nodes in a functional domain to the same f-structure, and thus allow

[2]Their proof strategy is quite straightforward. Following Wedekind (2014), they show that for an arbitrary context-free grammar an LFG grammar can be constructed whose c-structures simulate the context-free derivations and whose corresponding f-structures encode the terminal strings of those derivations. Grammars that encode the strings of two context-free grammars G_1 and G_2 can easily be composed so that the core computational problems for the combined LFG grammar are solvable if and only if the intersection $L(G_1) \cap L(G_2)$ is non-empty. This problem is known to be undecidable.

information to propagate up, down, and across the chain of nodes that relate to a single head. The grammar may implicitly impose a bound h on the height of a functional domain if vertical chains bottom out at lexical items or are terminated by nodes with other annotations. If the grammar does not establish an implicit bound on the height of the functional domains, one must be specified as an extragrammatical parameter. Given that such a bound has been determined, there is a simple transformation of a grammar G into a strongly equivalent LFG grammar $G^{\setminus\uparrow=\downarrow}$ that no longer contains $\uparrow=\downarrow$ annotations. The transformation is accomplished by recursively replacing a category annotated with $\uparrow=\downarrow$ in the right side of one rule by the right sides of all the rules expanding that category, and making the appropriate replacements of $\uparrow$ for $\downarrow$ to preserve the f-structure mappings. The effect of a simple case of this transformation is shown in (5).

(5)



The f-structure units are labeled here to illustrate how the size of the ϕ^{-1} node sets is reduced when $\uparrow=\downarrow$ annotations are eliminated. This simple grammar transformation makes it unnecessary to give further consideration to $\uparrow=\downarrow$ annotations.

Without further restriction, the recognition, emptiness, and generation problems are still undecidable for grammars with height-bounded functional domains. Thus, finite boundedness, a necessary condition for LCFRS equivalence, is also undecidable for these grammars.[3] WK observe that every LFG grammar G can be decomposed into two subgrammars, a *reentrancy-free kernel* $G_{\setminus\mathcal{R}}$ and an *atom-free kernel* $G_{\setminus\mathcal{A}}$, both with decidable recognition and emptiness problems, and they define the necessary and sufficient restrictions on the LFG grammars G by carefully regulating the interplay of the functional descriptions of these two grammars. The reentrancy-free kernel of G is formed by removing all reentrancies from

[3]This is because it is undecidable whether there is a valid derivation at all (undecidability of the emptiness problem).

its rules and lexical entries, leaving just function assignments and atomic-value annotations. The atom-free kernel is formed by removing all atomic-value annotations from its rules and lexical entries, so that only function assignments and reentrancies remain. We let $FD_{\setminus\mathcal{R}}$ and $FD_{\setminus\mathcal{A}}$ be the instantiated f-descriptions for derivations in $G_{\setminus\mathcal{R}}$ and $G_{\setminus\mathcal{A}}$ that correspond to a derivation in G with f-description FD.

Because d-boundedness is a necessary condition for LCFRS equivalence, WK first require the reentrancy-free kernel of $G^{\setminus\uparrow=\downarrow}$ to be d-bounded and they show that this is a decidable property. They then proceed to identify a minimally intrusive restriction that ensures that G conserves the ϕ mapping of its d-bounded reentrancy-free kernel. WK relate this restriction to another notion in the LFG literature, the concept of *nonconstructivity*. This has been discussed in the context of functional uncertainty and off-path constraints (e.g. Dalrymple et al., 1995b; Crouch et al., 2008), the idea that functional uncertainties pass information along separately motivated f-structure paths, and it is implicit in the fact that the Completeness Condition tests for the existence of independently specified grammatical functions. WK provide a technical formulation of this notion and propose it as a general condition on the operation of reentrancy annotations, a condition that is necessary to ensure LCFRS equivalence. They define nonconstructive reentrancies in terms of the interplay between the reentrancy-free and the atom-free kernel in the following way:

(6) G has **nonconstructive reentrancies** if and only if for every derivation in $G^{\setminus\uparrow=\downarrow}$ and any nodes n and n', if $\phi(n) = \phi(n')$ follows from $FD_{\setminus\mathcal{R}}$, then $\phi(n) = \phi(n')$ also follows from $FD_{\setminus\mathcal{A}}$.

This is a succinct formalization of the idea that reentrancies by themselves, without the support of function assignments, do not cause different nodes to project to the same f-structure. Thus a grammar is d-bounded if its reentrancy-free kernel is d-bounded and its reentrancies are nonconstructive. WK prove, for any LFG grammar G with only short reentrancies and a d-bounded reentrancy-free kernel, that it is decidable whether G has nonconstructive reentrancies.

The long reentrancies of the form $(\uparrow \text{F G}) = (\uparrow \text{H})$ that are found in descriptions of functional control, for example

(7) $(\uparrow \text{XCOMP SUBJ}) = (\uparrow \text{OBJ})$

appear to be just a minor enhancement to the short reentrancies in (4). However, WK demonstrate that core computational problems are undecidable for (1-)bounded grammars with long reentrancies. Thus, finite boundedness by itself is not a sufficient condition for LCFRS equivalence for grammars with long reentrancies that cannot be reduced to short ones.

WK observe that long control reentrancies such as (7) can always be shortened in derivations that meet the requirements of the Coherence Condition. They argue specifically that SUBJ is a governable function in an open (XCOMP) complement and therefore must be licensed by the complement's semantic form. These licensing semantic forms are always introduced by simple PRED equations associated with individual lexical entries, for example $(\uparrow \text{PRED}) = \text{'WALK}\langle(\uparrow \text{SUBJ})\rangle\text{'}$. Thus, $(\uparrow \text{PRED}) = \text{'WALK}\langle(\uparrow \text{SUBJ})\rangle\text{'}$ must instantiate to the equation $(\phi(n') \text{ PRED}) = \text{'WALK}\langle(\phi(n') \text{ SUBJ})\rangle\text{'}$ at some node n', and the functional description must also entail an equation $(\phi(n) \text{ XCOMP}) = \phi(n')$ that links the complement to a higher clause and is also available to shorten the control equation.

Even though it is easy to determine whether all long reentrancies can be shortened in any given derivation, it is not possible in general to construct an LCFRS that exactly simulates the set of all valid derivations of an LFG grammar. This is because it is also undecidable whether there are any derivations in which all long reentrancies can be shortened. Therefore, WK require the derivations to meet the stronger stipulation that the shortening equations ($(\phi(n) \text{ XCOMP}) = \phi(n')$) are entailed by the reentrancy-free kernel. This is formalized in (8).

(8) If FD contains $(\phi(n) \text{ F G}) = (\phi(n) \text{ H})$, then $FD_{\setminus\mathcal{R}}$ entails $(\phi(n) \text{ F}) = \phi(n')$ for some node n'.

Thus, just as LFG theory classifies as invalid derivations that do not satisfy the Completeness and Coherence Conditions (or that violate, in earlier specifications, the prohibition against nonbranching dominance chains), WK propose to remove from grammatical consideration derivations with recalcitrant control equations. This means that, as in the case of these previous conditions, some analyses will be excluded that otherwise appear to lie within scope of the normal derivational machinery of the LFG formalism. In all likelihood the derivations that this restriction eliminates would also fail to meet the Coherence Condition and thus no linguistically significant derivations will be lost.

The formal framework and result laid out by WK is summarized in (9).

(9) If G is an LFG grammar such that
  G includes only the reentrancies in (4)
  no more than one function assignment or $\uparrow = \downarrow$ annotation
    is attached to any category,
  G's functional domains are h-bounded,
  G's reentrancy-free kernel is d-bounded, and
  G's reentrancies are decidably nonconstructive
then G is equivalent to an LCFRS.[4]

[4]The LCFRS is constructed in two stages. In the first stage a $\uparrow = \downarrow$-free LFG grammar

WK also demonstrate that the emptiness and cyclic generation problems for these grammars are decidable and that sentences can be recognized in polynomial time.

3 The challenge of X-bar theory

WK suggest that grammars that meet the conditions in (9) and have the associated computational advantages are also suitable for linguistic description. This suggestion appears to be undermined by the principles of X-bar theory as they appear in the literature in various forms. These principles are generally assumed to constrain the organization of c-structure and the distribution of annotations that map from c-structure to f-structure (e.g. Bresnan, 2001; Dalrymple, 2001). X-bar prescriptions are typically given as meta phrase-structure rules that govern the expansion of a generic major category XP into an X′-labeled head which expands in turn to an X-labeled lexical head. These expansions branch, either recursively or iteratively, to other major categories annotated with function assignments $(\uparrow\ \text{G}) = \downarrow$ (for argument functions) or set-member annotations $(\downarrow \in (\uparrow\ \text{ADJ})$ for modifiers). They also branch to other categories with $\uparrow = \downarrow$ annotations (for coheads).

The problem for finite boundedness is that X-bar compliant rule systems appear to allow for arbitrary repetitions of nodes, particularly complements and coheads, that map to the same f-structure. Schematically, the possibilities for specifiers, complements, and coheads are illustrated in (10). (The infix comma notation is a conventional way of abbreviating the fact that this scheme is agnostic as to the linear order of constituents, and any specific categories may or may not appear in the particular rules that instantiate this general scheme.)

(10) a. $$\text{XP} \rightarrow \underset{\substack{\uparrow=\downarrow \\ \\ \textit{head}}}{\text{X}'}\ ,\ \underset{\substack{(\uparrow\ \text{D})=\downarrow \\ \\ \textit{spec}}}{\text{LP}\,|\,\text{FP}}$$

b. $$\text{X}' \rightarrow \underset{\substack{\uparrow=\downarrow \\ \\ \textit{head}}}{\text{X}'}\ ,\ \underset{\substack{\textit{comp} \\ \textit{cohead}}}{\left\{\begin{matrix} \text{LP} \\ (\uparrow\ \text{G})=\downarrow \\ \text{FP} \\ \uparrow=\downarrow \end{matrix}\right\}}$$

$G^{\backslash\uparrow=\downarrow}$ is created by eliminating the h-bounded $\uparrow=\downarrow$-annotated categories in favor of equivalent collections of flattened LFG rules. The second stage of the construction produces LCFRS rules for $G^{\backslash\uparrow=\downarrow}$. It hypothesizes finite sequences of $G^{\backslash\uparrow=\downarrow}$ rules that might expand the categories realizing a d-bounded zipper, and it builds an LCFRS rule that models the well-formedness conditions and the minimal f-structure for each such sequence.

c. $X' \rightarrow \underset{\substack{\uparrow = \downarrow \\ \textit{head}}}{X} \quad , \quad \underset{\substack{(\uparrow \text{ G}) = \downarrow \\ \textit{comp}}}{LP}$

In this particular version of the meta-grammatical scheme, the lexical or functional categories LP or FP in (10a) assign a discourse function (D) to the specifier. The recursive expansion in (10b) derives complements (labeled as governable grammatical functions G) and coheads, and the recursion terminates at the lexical head in (10c).[5]

We focus our attention on the X′ recursion in (10b). This permits the object of a VP, for example, to be realized discontinuously by any number of NP's annotated with $(\uparrow \text{ OBJ}) = \downarrow$, and all of those nodes would contribute features to the same f-structure. Thus, this general rule schema immediately licenses derivations that are not finitely bounded. Or, to put it in more traditional terms, the X-bar framework admits of grammatical rules that allow for an unbounded degree of discontinuity (Chomsky, 1953). This scheme also allows for functional domains of unbounded height, to derive arbitrary numbers of coheads in addition to a possible head. Grammars formulated according to this specification clearly fail to meet the tractability requirements as summarized in (9).

As theoretically appealing as it may be, this configuration may not be descriptively accurate for real languages. Real languages may have substantially less potential for repetition, and rules of this type may substantially overgenerate beyond what should be implemented in descriptively adequate grammars. Indeed, we suspect that the unrestricted X-bar schema does overgenerate, and that functional units in real languages are not arbitrarily decomposable. Individual lexical predicates subcategorize for a limited number of governable grammatical functions, there are a limited number of morphosyntactic and cohead features to be expressed with limited redundancy, and there may also be categorial cooccurrence restrictions that limit, for example, which subconstituents internal to an NP can surface independently (e.g. perhaps agreement markers and determiners must always appear together). And of course, because semantic forms are instantiated in LFG derivations, there can be at most one lexical head among the phrases that realize any grammatical function. We thus conjecture that for each language the X-bar scheme is constrained by two extragrammatical parameters: the maximum degree of discontinuity d and the maximum number c of coheads in a functional domain.

The constants d and c provide an upper bound on the height of the X′ recursion in (10b), since there is also a bound on a language-dependent parameter g, the number of grammatical functions (discourse functions and

[5]Some presentations make use of an iterative Kleene-star format instead of the recursive, binary-branching formulation here. Both formulations admit of unbounded repetition and are thus essentially the same with respect to the formal issues we are concerned with here.

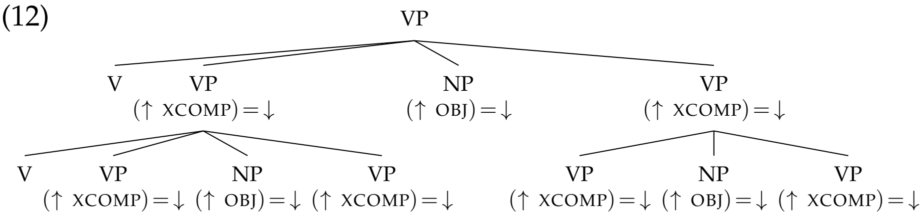
governable functions) that can appear together in any subtree.[6] The maximum height h is given by the formula $h = c + dg + 1$, with the additional 1 accounting for the lexical X expansion in (10b). If a subderivation for XP is higher than h, then the d bound must be exceeded for at least one grammatical function or the c bound must be violated. For a grammar that respects this height limit the algorithm for eliminating $\uparrow = \downarrow$ annotations from G will terminate in a grammar $G^{\setminus\uparrow=\downarrow}$ with a finite number of rules.

In the simplest case the succinct recursion in (10) would be converted to flattened rules that explicitly limit the number of repetitions of any one function assignment. For $d = 2$, for example, such a grammar might contain the V′ expansion rule (11) that maps only two nodes to XCOMP and one node to OBJ.

(11) $$\begin{array}{llllll} \mathrm{V}' & \rightarrow & (\mathrm{V}) & \mathrm{VP} & \mathrm{NP} & \mathrm{VP} \\ & & & (\uparrow \textsc{xcomp}) = \downarrow & (\uparrow \textsc{obj}) = \downarrow & (\uparrow \textsc{xcomp}) = \downarrow \end{array}$$

It is not difficult to see how this rule relates to the recursive X-bar specification, and the connection to an iterative Kleene-star X-bar schema would be even more obvious: a grammar's *-marked generic rules would be reduced to particular sequences whose length is restricted by h. The fragmentary V′ expansion shown in (12) demonstrates, however, that restricting the length of individual rules is by itself not sufficient to ensure that discontinuity is globally bounded.

(12)



Although only two VP nodes map to the same f-structure at each level of recursion, the overall effect of the parallel subderivations (with zipper unification) is that the four discontinuous VP nodes at the bottom map to the

[6] The grammatical functions attested in any language are drawn from a finite universal set GF, and g cannot be greater than $|\textsc{gf}|$. The number of governable functions included in GF can be determined for any particular language by inspecting the subcategorization frames in its lexicon. For the broad-coverage, commercial-grade Pargram grammar of English (approximately 25,000 lexical entries) there are 10 different governable functions, and for the German Pargram grammar there are 13. But a tighter bound on g comes from the fact that no word in either lexicon governs more than four functions, and very few words allow even that many (in English only the word *bet*). The inventory of lexical subcategorization frames further restricts the cooccuring combinations of governable functions. For the English and German grammars the number of combinations (32 and 41 respectively) is much less than the theoretical maximum of all up-to-four combinations of the attested set of governable functions.

same XCOMP XCOMP f-structure. While it is possible in principle to convert an X-bar compliant grammar (either recursive or iterative) to one that is descriptively accurate for a language with a global bound on the degree of discontinuity, the resulting grammar will have an elaboration of rules and a refinement of features and categories that will likely be convoluted and opaque, and its relation to the explanatory generalizations of X-bar theory will be obscure.

We propose to address this discrepancy between explanatory and descriptive adequacy, for a language with a bounded degree of discontinuity and a bounded number of coheads, in a direct and brute-force way. Such a language is described by an X-bar compliant grammar together with separate d and c values that bound its degree of discontinuity and the number of coheads under a single XP. If a grammatical function can be realized by no more than d separate XP's one of which contains a lexical head, and each XP can have no more than c coheads, then the number of nodes that can map to a single f-structure is bounded by $dc + 1$. We simply declare as invalid and disallow any derivation if $|\phi^{-1}(f)|$ for any f exceeds that number.[7] This is a further filtering condition on derivations, along the lines of Completeness and Coherence and the original nonbranching dominance prohibition of Kaplan and Bresnan (1982). Once the c-structure and the minimal solution of the f-description have been constructed, derivations are discarded if their structure-function mapping exceeds this bound.

Derivations that survive the *d*-*c* restriction also satisfy the bounding requirements as stated in (9). This at least removes the specific challenge to tractability that comes with the unbounded repetition of arguments and coheads implied by the principles of X-bar theory. If grammars for natural language also meet the constraints on notation and reentrancies in (9), as WK conjecture, then it is possible to construct linear context-free rewriting systems for such grammars that simulate all and only their valid derivations.[8]

4 Modifiers: Height and discontinuity

Modifiers are represented as sets in f-structure precisely because they are not selected by particular predicates and because there are no natural limits on how many may appear. They therefore pose a different kind of challenge to the bounding conditions necessary for LCFRS equivalence. One

[7]Note that $dc + 1$ is also an upper bound for a single grammatical function. This architecture is flexible enough to account for much more fine-grained distinctions where, for example, the major categories or grammatical functions differ in their degree of discontinuity and/or their number of coheads.

[8]In constructing an LCFRS for derivations with extragrammatical bounds, only reentrancies in the bounded derivations must be checked for nonconstructivity. This can be decided along the lines of the shrinking argument of WK.

aspect of the problem is illustrated by the Kleene-starred adjunct phrases of Bresnan's (2001) alternative X-bar schema in (13).

(13) $$\begin{array}{lllll} \mathrm{XP} & \rightarrow & \mathrm{X}' & , & \mathrm{YP}^{*} \\ & & \uparrow = \downarrow & & \downarrow \in (\uparrow \text{ ADJ}) \end{array}$$

These rules can be normalized to rules of the recursive form by translating the Kleene-starred adjunct phrases into right-linear expansions. The translation is done by replacing the iteration of the phrasal adjunct in (13) by a new optional trivially-annotated category YP_{ADJ}, and by introducing the rule (14) to properly expand that category.

(14) $$\begin{array}{llll} \mathrm{YP}_{\text{ADJ}} & \rightarrow & \mathrm{YP} & (\,\mathrm{YP}_{\text{ADJ}}\,) \\ & & \downarrow \in (\uparrow \text{ ADJ}) & \uparrow = \downarrow \end{array}$$

Immediately we see that there is no apparent bound on the height of the functional domain and thus no way of eliminating these trivial annotations. But, as it turns out, the LCFRS construction does not require the removal of trivials in this particular configuration. This is because these YP_{ADJ} nodes and their subtrees are inert with respect to zipper interactions. Therefore these nodes can be ignored when determining the height of the functional domain, the degree of discontinuity, and the finite bound as given in (3).

The expansion of a category is syntactically inert if the zippers within its subtrees in all possible derivations cannot interact with the zippers outside. This is certainly true of the YP expansion in (14), since membership annotations, unlike function assignments, form barriers that prevent daughter attributes and values from escaping to higher levels. The expansions of the trivially-annotated YP_{ADJ} categories are also inert because they dominate only inert category expansions. Thus the YP_{ADJ} nodes need not enter into the calculation of boundedness and their $\uparrow = \downarrow$ annotations need not be eliminated. We mark that fact simply by replacing those equalities with a variant $\uparrow \doteq \downarrow$ that is opaque to the trivial-elimination procedure. It can be carried along by the LCFRS translation algorithm and interpreted as a node identity only during f-structure construction.

Internal adjuncts of discontinuous NPs are another instance of inertness. A Latin example taken from Haug (2017) is shown in (15).

(15) S

- NP $(\uparrow \textsc{subj}) = \downarrow$
  - N $\uparrow = \downarrow$
    - *Maximilianus* $(\uparrow \textsc{pred}) = \text{‘}\textsc{Max}\text{’}$
- NP $(\uparrow \textsc{obj}) = \downarrow$
  - AdjP $\downarrow \in (\uparrow \textsc{adj})$
    - Adj $\uparrow = \downarrow$
      - *bonum* $(\uparrow \textsc{pred}) = \text{‘}\textsc{good}\text{’}$
- I $\uparrow = \downarrow$
  - *trusit* $(\uparrow \textsc{pred}) = \text{‘}\textsc{push}\langle\textsc{subj}, \textsc{obj}\rangle\text{’}$
- NP $(\uparrow \textsc{obj}) = \downarrow$
  - N $\uparrow = \downarrow$
    - *Fredericum* $(\uparrow \textsc{pred}) = \text{‘}\textsc{Fred}\text{’}$

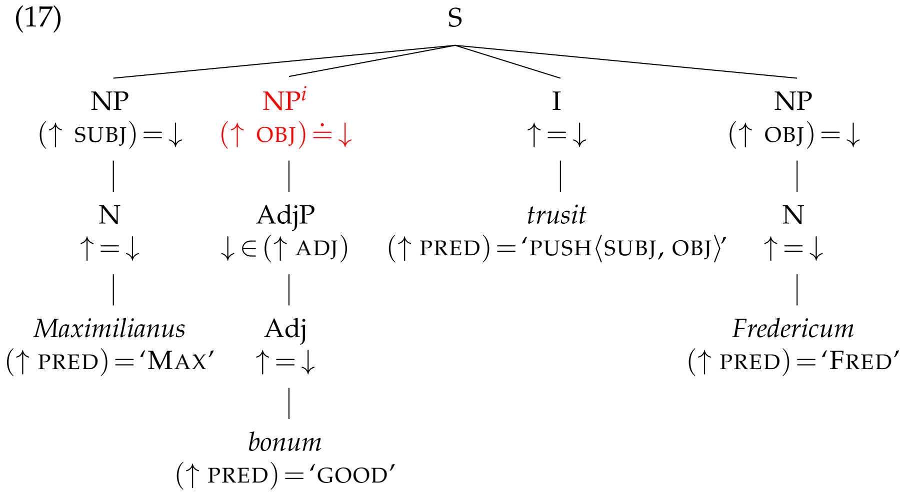
Haug (2017) investigates the degree of discontinuity of Latin based on the nonprojective dependencies occurring in several dependency treebanks and argues that there is no principled bound on Latin discontinuities. In his examples, however, the adjunct NP expansions that seem to lead to this conclusion are—as in (15)—inert with respect to zippers, because no information can escape from their AdjP daughters. This might result in an unbounded number of constituents that map to the same f-structure, but crucially only the f-structures of a bounded number of them can interact and thus give rise to zipper dependencies.

In this situation the adjunct expansion of the NP can be marked as inert, as indicated by the superscript *i* in (16),

(16) $\text{NP}^{i} \;\rightarrow\; \underset{\downarrow \in (\uparrow \textsc{adj})}{\text{AdjP}}$

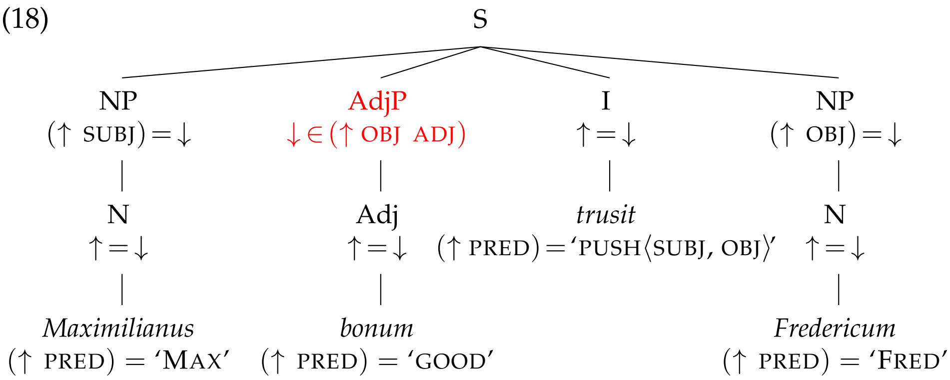
and the NP occurrence in the S rule can be replaced by NP^{i} with a variant annotation $(\uparrow \textsc{obj}) \doteq \downarrow$ that explicitly indicates that this function assignment does not give rise to discontinuities that prevent the construction of an equivalent LCFRS. Instead of (15) we thus obtain the annotated c-structure in (17).

(17)



Alternatively, such examples can be reanalyzed as in (18) so that inert nodes are no longer present in the c-structure, as suggested by Snijders (2016). This also ensures that the number of nodes that map to the same f-structure is finitely bounded.

(18)



There may be no limit on the number of discontinuous nominal adjuncts in Latin, but their inertness means that a *d*-bounded LFG can still account for them, contrary to Haug's supposition.

Haug also recalls Johnson's (1986) observation that the Bresnan et al. (1982) analysis of Dutch cross-serial dependencies does not extend to intransitive verb complexes without violating the nonbranching dominance constraint. In response, Zaenen and Kaplan (1995) covered both transitive and intransitive examples with a functional uncertainty solution that shortens the nonbranching chains. Haug notes that the newer formulation is less transparent, linguistically less attractive, and still less often cited than the original 1982 account. Because of this and because of apparently similar examples in Latin, Haug is willing to give up the prohibition against nonbranching dominance chains and its guarantee that recognition is decidable for unrestricted LFG grammars, in favor of simpler nonbranching specifications. However, the simpler specifications can be implemented in the bounded-grammar framework, and WK have demonstrated that the nonbranching dominance constraint is not needed for recognition decidability and can be omitted from the restricted formalism.

5 Conclusion

Wedekind and Kaplan (to appear) have shown that there is a mildly context-sensitive grammar, an LCFRS, for every LFG grammar with the properties set forth in (9). The grammars in this class are mathematically and computationally well behaved: for languages that can be described by such grammars, the recognition problem is decidable and tractable (even without the prohibition against nonbranching dominance chains), and the emptiness and cyclic generation problems are decidable. The explanatory principles of X-bar theory admit grammars that appear not to meet the bounding conditions that WK have identified and thus seem to lie outside of the tractable class.

This may be an unintended consequence of the simple way in which the X-bar principles are typically laid out, and may in fact result in the overgeneration of strings and structures. We have provided a formal definition of a traditional linguistic notion, the degree of discontinuity for grammatical functions of a language, in terms of the c-structure to f-structure mapping of LFG, the ϕ projection. If the degree of discontinuity and the number of coheads for a language are bounded, then the derivations of an otherwise X-bar compliant grammar can be restricted to meet the conditions necessary for tractability. Notably, these bounds generally do not apply to modifier repetition because modifiers are typically isolated from their environments by set membership annotations. Thus, we suggest adding to the meta-theory of LFG these bounding requirements as a stronger replacement for the previous prohibition of nonbranching dominance chains. We conjecture that this will enable explanatory grammars that are not only descriptively accurate but also computationally tractable.

References

Berwick, Robert C. 1982. Computational complexity and Lexical-Functional Grammar. *American Journal of Computational Linguistics*, 8(3–4):97–109.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell Publishers, Oxford.

Bresnan, Joan, Ronald M. Kaplan, Stanley Peters, and Annie Zaenen. 1982. Cross-serial dependencies in Dutch. *Linguistic Inquiry*, 13(4):613–635.

Chomsky, Noam. 1953. Systems of syntactic analysis. *Journal of Symbolic Logic*, 18(3):242–256.

Crouch, Dick, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell, III, and Paula S. Newman. 2008. *XLE Documentation*. Palo Alto Research Center, Palo Alto, CA.

Dalrymple, Mary. 2001. *Lexical-Functional Grammar*. Academic Press, New York.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell, III, and Annie Zaenen, editors. 1995a. *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, Stanford University.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell, III, and Annie Zaenen. 1995b. Non-local dependencies. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell, III, and Annie Zaenen, editors, *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, Stanford University, pages 131–135.

Haug, Dag. 2017. Syntactic discontinuities in Latin. In Victoria Rosén and Koenraad De Smedt, editors, *The Very Model of a Modern Linguist: In Honor of Helge Dyvik*. Bergen Language and Linguistic Studies, Bergen, pages 75–96.

Johnson, Mark. 1986. The LFG treatment of discontinuity and the double infinitive construction in Dutch. In *Proceedings of the Fifth West Coast Conference on Formal Linguistics*, pages 102–118, Stanford Linguistics Association, Stanford.

Kallmeyer, Laura. 2013. Linear context-free rewriting systems. *Language and Linguistics Compass*, 7(1):22–38.

Kaplan, Ronald M. and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Mass., pages 173–281. Reprinted in Dalrymple et al. (1995a, 98–111).

Maxwell, John T., III and Ronald M. Kaplan. 1996. Unification parsers that automatically take advantage of context freeness. In *Proceedings of the LFG'96 Conference*, CSLI publications, Stanford University.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science*, 88(2):191–229.

Seki, Hiroyuki, Ryuichi Nakanishi, Yuichi Kaji, Sachiko Ando, and Tadao Kasami. 1993. Parallel multiple context-free grammars, finite-state translation systems, and polynomial-time recognizable subclasses of Lexical-Functional Grammars. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 130–139, Association for Computational Linguistics, Columbus, OH.

Snijders, Liselotte. 2016. An LFG account of discontinuous nominal expressions. In *Proceedings of the Workshop on Discontinuous Structures in Natural Language Processing*, pages 1–11, Association for Computational Linguistics, San Diego, California.

Wedekind, Jürgen. 2014. On the universal generation problem for unification grammars. *Computational Linguistics*, 40(3):533–538.

Wedekind, Jürgen and Ronald M. Kaplan. 2012. LFG generation by grammar specialization. *Computational Linguistics*, 38(4):867–915.

Wedekind, Jürgen and Ronald M. Kaplan. To appear. Tractable Lexical-Functional Grammar. *Computational Linguistics*.

Zaenen, Annie and Ronald M. Kaplan. 1995. Formal devices for linguistic generalizations: West Germanic word order in LFG. In Jennifer Cole, Georgia M. Green, and Jerry L. Morgan, editors, *Linguistics and Computation*. CSLI Publications, Stanford, pages 3–27. Reprinted in Dalrymple et al. (1995a, 215–239).

Pronominal Possessors and Syntactic Functions in the Hungarian Possessive Noun Phrase

Tibor Laczkó
Károli Gáspár University of the Reformed Church in Hungary

György Rákosi
University of Debrecen

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 149–169

http://csli-publications.stanford.edu/LFG/2019

Keywords: Hungarian, possessive noun phrase, binding, reciprocal pronoun, reflexive pronoun

Laczkó, Tibor, & Rákosi, György. 2019. Pronominal Possessors and Syntactic Functions in the Hungarian Possessive Noun Phrase. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 149–169. Stanford, CA: CSLI Publications.

Abstract

In this paper we develop an LFG analysis of the binding relations of Hungarian anaphors when they occur within possessive DPs. The reflexive is subject to the Minimal Complete Nucleus Condition, and the reciprocal is subject to the Minimal Finite Domain Condition. When either the reflexive or the reciprocal pronoun occurs within a possessive DP, neither of them can be anaphorically bound from outside if this DP contains the definite article (Rákosi 2017, to appear). Our analysis has two crucial aspects. On the one hand, we introduce a new feature: "binding domain delimiter" associated with the lexical form of the definite article. We use this feature as a negative off-path constraint in modelling the relevant binding relations. On the other hand, following Laczkó (2004, 2009), we assume that within Hungarian possessive DPs there are two [–r] grammatical functions available to arguments of complex event nominals: POSS and SUBJ. Both can be overtly realized by either the nominative or the dative possessor constituent, and, in addition, SUBJ can also be PRO. Thus, we create a DP-internal antecedent for the anaphors in a principled manner, which, in turn, can be controlled from outside the DP.

1. Introduction

The Hungarian possessive noun phrase can host a wide range of pronominal possessors: personal pronouns, reflexives, as well as the reciprocal anaphor are each licensed as possessors. Each of these pronominal possessors can form a referential dependency with a clause-mate antecedent.[1]

This paper presents an in-depth LFG analysis of the syntax of anaphoric possessor strategies in Hungarian, and it makes two fundamental claims. First, following Rákosi (2017, to appear), we show that the definite article plays a crucial blocking role, inasmuch as bound variable readings between possessor anaphors and clause-mate antecedents are licensed only in the absence of the article. Second, we argue that the proper LFG treatment of these anaphoric dependencies necessitates the postulation of a SUBJ function internal to the possessive noun phrase that co-exists with POSS in the case of nominalization (Laczkó 2004, 2009).

The structure of the paper is as follows. In section 2, we present an overview of the major anaphoric possessor strategies in Hungarian on the basis of Rákosi (2017, to appear), paying special attention to the distribution of the definite article. We also make some remarks on the binding domains that generally characterise reflexive and reciprocal anaphors in Hungarian.

[1] By *possessive noun phrase*, we mean the NP/DP that has a POSS argument within its own f-structure (*the girl's hand*), and we use the term *possessor* to mean the NP/DP that fulfils the POSS GF (*the girl's* in *the girl's hand*).

We present an LFG analysis of these data in section 3, and conclude the paper with a summary in section 4.

2. The definite article and anaphoric possessors

2.1. A puzzling distribution of the article

The distribution of the definite article across the different Hungarian pronominal possessor constructions appears to be puzzling at first: the article is optional if the possessor is *pro*-dropped (1a), it is obligatory if the personal pronoun possessor is overt (1b), it is also obligatory if the possessor is a reflexive (1c), but it is barely an option if the possessor is the reciprocal anaphor (1d).

(1) *A tanár-ok*$_i$ *ismerték*
the teacher-PL knew.3PL
'The teachers$_i$ knew… '

a. [$_{DP}$ (*a*) *határ-a-i-k-at*].
the limit-POSS-PL-3PL-ACC
' …their$_{i/j}$ limits.'

b. [$_{DP}$ *(*az*) *ő*$_{i/j}$ (*kis*) *határ-a-i-k-at*].
the (s)he (little) limit-POSS-PL-3PL-ACC
' …their$_{i/j}$ (little) limits.'

c. [$_{DP}$ *(*a*) *maguk*$_{i/*j}$ *határ-a-i-t*].
the themselves limit-POSS-PL-ACC
' …their$_{i/*j}$ own limits.'

d. [$_{DP}$ ($^{*/??}$*az*) *egymás*$_{i/*j}$ *határ-a-i-t*].
the each_other $_{i/*j}$ limit-POSS-PL-ACC
' …each other's limits.'

Pronominal possessors agree with the possessum in Hungarian, and the morphology on the inflected head noun shows an intricate complexity. In (1a), for example, the *possessedness* morpheme *-a-* follows the head, then the plural marker *-i-* is used to pluralize the possessum. It is followed by the 3PL agreement marker *-k-*, which incorporates the 3PL pronominal possessor; and the accusative case marker *-t* comes last in the sequence. Since this morphology identifies pronominal possessors, these are regularly dropped, as in (1a). The overt possessor pronouns in (1b) shows no number agreement with the inflected 3PL head in third person, and it is spelt out as a 3SG possessor (this pattern is known as *anti-agreement* in Hungarian grammars). The reflexive (1c) and the reciprocal (1d) possessors show no agreement with the head.[2]

The most puzzling fact about the distribution of the definite article across the constructions in (1) is that the reflexive possessor (1c) and the pronominal possessor (1b) pattern up in requiring the definite article, whereas the reciprocal possessor cannot take it. Rákosi (2017, to appear) argues that this intricate pattern is in fact predictable if we assume that the definite article plays a role in delimiting the respective binding domains. The pertinent literature makes two claims that we may utilize as vantage points in spelling out an adequate account.

First, both É. Kiss (1987: 197-202) and Marácz (1989: 391-398) argue that the Hungarian possessive noun phrase is a binding domain. This, É. Kiss notes, renders the reflexive possessor strategy in (1c) a "marked pattern", placing the reflexive possessor "outside of the domain of binding theory, into the periphery of grammar" (1987: 198). As we briefly show below, the reflexive here is indeed an exempt anaphor in the sense of Pollard & Sag (1992), and it has logophoric properties. It is "marked" in the sense that logophoric pronouns have a marked character: they always require a supporting discourse context wherein the perspective holder that can be construed as an antecedent is available.[3] The reciprocal possessor does not need such a supportive discourse context, all it requires in the usual case is an available antecedent within the clause. Second, Marácz (1989) notes the lack of the article in the case of the reciprocal (1d), which leads him to conclude that for reciprocals, the embedding clause acts as a binding domain. For the construction represented by (1d), we will make the same assumption.[4]

2.2. Two notes on the binding domains

Since our goal is a unified analysis of reflexive and reciprocal anaphors (strictly distinguishing these in the lexicon from the corresponding logophoric entries, which we treat as exempt elements), it is useful to add two comments on the binding domains that they are constrained by. Note, first of all, that both anaphors figure in predicative PPs taking the clausal

[2] For a detailed LFG-specific discussion of the grammar of the Hungarian possessive noun phrase, see Laczkó (1995).
[3] The lack of the definite article with reflexive possessors leads to ungrammaticality, and its presence still leaves the reflexive possessor here a less frequent alternative to the *pro*-drop construction in (1a), other things being equal.
[4] Marácz (1989) assumes that the definite article is never compatible with reciprocal possessors. We point out below that this assumption is not warranted, as there are cases when a reciprocal possessor is compatible with the definite article. In nominalizations, where the search for an antecedent may terminate inside the possessive noun phrase, the article becomes an option.

subject as their antecedents. This entails that the binding domain is not the coargument domain for either.[5]

(2) a. *A fiúk látták ez-t maguk mellett / *melletük.*
the boys saw.3PL this-ACC themselves next.to next.to.3PL
'The boys saw this next to them.'

b. *A fiúk láttak valami-t egymás mellett.*
the boys saw.3PL something-ACC each.other next.to
'The boys saw something next to each other.'

An interesting contrast emerges between reflexive and reciprocal anaphors in infinitival constructions. Compare the following two sentences:

(3) a. *A fiúk*$_i$ *látták a lányok-at*$_k$ *lerajzol-ni maguk-at*$_{*i/k}$.
the boys saw.3PL the girls-ACC draw-INF themselves-ACC
'The boys saw the girls draw (a picture of) themselves.'

b. *A fiúk*$_i$ *látták a lányok-at*$_k$ *lerajzol-ni egymás-t*$_{i/k}$.
the boys saw.3PL the girls-ACC draw-INF each.other-ACC
'The boys saw the girls draw (a picture of) each other.'

If the reflexive is the object argument of the infinitive, it has to be bound by the subject of the infinitive. Since (3) is a raising construction, the infinitival subject is controlled by the matrix object.[6] Consequently, the reflexive anaphor picks the girls in (3a), and the matrix subject is not a potential antecedent. But for the reciprocal, it is: the anaphor in (3b) can either be about the girls or the boys. We conclude therefore that reflexive anaphors are subject to the Minimal Complete Nucleus Condition in Hungarian, but the reciprocal can find an antecedent within the Minimal Finite Domain.[7]

Note nevertheless that this difference, by itself, does not account for the observations we made in 2.1 above. Most importantly, it makes no predictions with respect to the observed distribution of the definite article in possessive phrases. In the next subsection, we now turn to a more detailed

[5] Whereas the default choice is the pronoun in English in such configurations (see the translation of (2a)), the reflexive is the usual and often the only grammatical choice in Hungarian. See Rákosi (2010) for an LFG-specific discussion of these so-called *snake sentences* in Hungarian.

[6] This is an ordinary case of a "subject-to-object raising" construction. The infinitival constituent has the customary XCOMP function, and its covert subject is functionally controlled by the (formal) object of the matrix verb. Thus, the "immediate" binder of the reflexive object in the infinitival construction is the covert subject.

[7] See Dalrymple (2001) for an overview and a definition of the binding domains that are employed in LFG grammars.

discussion of this distribution and its relevance in licensing referential dependencies between anaphoric possessors and their antecedents.

2.3. More about anaphoric possessors

A recent line of research has found a strong typological correlation between the availability of *dedicated possessive reflexives* and the way languages code definiteness (see Reuland 2007, 2011, Despić 2011, 2015, Marelj 2011). Such dedicated possessive reflexives are only available in languages which do not employ prenominal definite articles (i.e., only in languages with postnominal definiteness marking or with no definiteness marking at all, see Despić 2015: 203 for a representative list). Latin and Italian form a minimal pair in this respect: Latin has no definite article and it has the dedicated possessive reflexive *suus*, but Italian has a definite article and it has only an English-type pronominal possessor. Compare (4a) and (4b) below for illustration. The Latin possessive phrase does not act as a binding domain, which results in the classical complementarity between the two types of pronominal possessors, but the Italian possessive phrase, armoured with the definite article, is a binding domain. As a result, Italian has only one type of possessive pronoun, and the contrast that Latin entertains has been lost.

(4) a. Latin (Bertocchi & Casadio: 1980, 26)

Ioannes$_{i}$	*sororem*	*suam*$_{i/*k}$ /	*eius*$_{*i/k}$	*vidit*.
Ioannes	sister.ACC	self's	his	saw

'Ioannes saw his sister.'

b. Italian (Reuland 2011: 168)

Gianni$_{i}$	*ama*	*le*	*sue*$_{i/k}$	*due*	*machine*.
Gianni	loves	the	his	two	cars

'Gianni loves his two cars.'

Rákosi (2017, to appear) argues that Hungarian instantiates, as it were, both of these universal scenarios. The reciprocal possessor can be a true anaphor bound by the clausal subject in the absence of the definite article (1d), and when the definite article is there (1a-c), the dependency between the anaphoric possessor and the main-clause antecedent is essentially a long distance dependency.

This is straightforward for personal pronoun possessors, which, as expected, should co-occur with the definite article if the article indeed spells out the left edge of a binding domain.[8] It is reflexive possessors that do not

[8] In fact, overt personal pronoun possessors always require the presence of the definite article in Hungarian, irrespective of whether they have a clause-mate antecedent or not. When they do have a clause-mate antecedent, the usual strategy is

appear to be well-behaved at first, since they require the presence of the definite article (1c). In fact, as Rákosi (to appear) argues in detail, reflexive possessors in Hungarian are discourse sensitive, exempt anaphors. This is most obvious when they do not have a clause-mate antecedent, as in the following example below (source: *Hungarian National Corpus*).

(5)	*Elég*	*nagy*	*így is*	*a*	*magam*	*terh-e*!
	quite	big	even so	the	myself	burden-POSS.3SG
	'My own burden is quite big even so.'					

We will consequently treat these reflexive possessors as special, discourse sensitive pronominal elements, which may not even have linguistically expressed antecedents at all.

Reciprocal possessors, on the other hand, are well-behaving anaphors, and the definite article has a complex distribution in their case which is fully compatible with this assumption. Consider the following sentences for illustration, each of which represents a different reciprocal possessor construction.

---

to *pro*-drop the possessor, and spelling it out is a marked option in most contexts. The insertion of the speaker-oriented modifier *kis* 'little' is one strategy that makes the use of an overt pronoun more natural in the presence of clause-mate antecedents, that is why we added this adjective in (1b).

The definite article can sometimes be absent if the pronominal possessor is *pro*-dropped. The conditions licensing such article-drop are complex, but it is best if the possessive phrase has a salient referent in the discourse. Compare these two examples:

(i)	*Szeretem*	#(*az*)	*ablak-om-at.*
	love.1SG	the	window-POSS.1SG-ACC
	'I love my window.'		

(ii)	*Szeretem*	(*az*)	*anyá-m-at.*
	love.1SG	the	mother-POSS.1SG-ACC
	'I love my mother.'		

Unlike in Italian, the omission of the article is not determined solely by choice of the noun head (though this is a primary factor), but it may be subject to contextual parameters. We do not discuss these here, as our main concern in this paper is a study of reflexive and reciprocal possessors. But note that the article is always grammatical with either overt or *pro*-dropped pronoun possessors, and that it can be sometimes omitted in the latter case is not relevant for our analysis to be presented in Section 3.

(6) a. *Jól ismerjük* [DP (*/??*az*) *egymás* *baj-á-t*].
well know.1PL the each_other problem-POSS.3SG-ACC
'We know each other's problems well.'

b. *Egymás-nak jól ismerjük* [DP *(*a*) *baj-á-t*].
each_other-DAT well know.1PL the problem-POSS.3SG-ACC
'We know each other's problems well.'

c. *A fiúk*$_i$ *díjazzák* [DP (*az*) *egymás*$_i$ *lefest-és-é-t*].
the boys appreciate.3PL the each_other paint-DEV-POSS.3SG-ACC
'The boys appreciate the painting of each other.'

(6a) represents a canonical transitive structure where the article is not acceptable, as we have also seen for (1d) above. When the possessor is extracted (and receives dative case), the spellout of the article is compulsory (6b). Notice that in this case the reciprocal is outside of the possessive phrase, and its local antecedent is the (*pro*-dropped) 1SG subject. Finally, (6c) contains a possessive phrase where the possessum is a deverbal nominal. At least when the understood subject of this nominalization is coreferential with the matrix subject, the definite article becomes optional for most native speakers, see Rákosi (to appear) for a discussion of pertinent questionnaire data. In this interpretation (when the boys appreciate their own painting of each other) the reciprocal has a syntactically active potential antecedent within the possessive nominalization. It forms an important part of our analysis presented in section 3 that nominalizations may include a SUBJ function internal to the possessive noun phrase. What we have shown in this section is that the definite article is indeed a binding domain delimiter in Hungarian possessive constructions, and this must be captured by any adequate analysis of the data we have surveyed here.

3. An LFG-account

In this section, we set out to develop an analysis for the following empirical generalizations, based on the data and the relevant discussions is section 2.

The primary Hungarian reflexive pronoun can be used either anaphorically or logophorically. In the former case, its binding domain is the minimal constituent containing a subject, i.e. the Minimal Complete Nucleus Condition applies to it. As should be clear from the foregoing discussion, it is the behaviour of reciprocal pronouns that poses a much greater challenge for a theoretical approach, so in this section our main focus will be the development of an adequate account of these reciprocal phenomena. However, at the end of the section we will also show that the analysis of the binding relations of the reflexive pronoun when it occurs in possessive DPs headed by a complex event nominal can be made more principled (and uniform) if it is cast in the general formal approach developed for reciprocals.

Reciprocal pronouns have been shown to be subject to the Minimal Finite Domain Condition, see the crucial example in (3b), and compare it with (3a) containing a reflexive pronoun. This condition allows reciprocal possessors to search for antecedents either inside or outside of the possessive phrase. However, it is an overall constraint on anaphoric dependencies involving pronominal possessors that the search for the antecedent cannot pass the definite article in the DP cap of the possessive phrase, see the crucial example in (6a), repeated here for convenience. It contains a reciprocal pronoun and an ordinary (nonderived) noun head in the possessive DP. The reciprocal is bound by the pro-dropped subject of the matrix verb The presence of the definite article blocks binding from outside the DP, and, given that there is no potential binder within the DP, the sentence is ungrammatical.[9]

(6a) *Jól ismerjük* [$_{\text{DP}}$ ($^{*/??}$*az*) *egymás* *baj-á-t*].
well know.1PL the each_other problem-POSS.3SG-ACC
'We know each other's problems well.'

The situation is complicated by the fact that the same construction type is fully acceptable, if the noun head in the possessive DP is a derived (complex event) nominal, see (6c), repeated below for convenience. If there is no definite article in the DP, the matrix subject can bind the reciprocal in the usual way, as in (6a). The presence of the article and the possible coreference of the reciprocal and the matrix subject requires a special treatment.

(6c) *A fiúk*$_i$ *díjazzák* [$_{\text{DP}}$ (*az*) *egymás*$_i$ *lefest-és-é-t*].
the boys appreciate.3PL the each_other paint-DEV-POSS.3SG-ACC
'The boys appreciate the painting of each other.'

Our approach then needs to achieve two goals. On the one hand, it has to formally encode the fact that the definite article, as a rule, marks the boundary of a binding domain for reciprocals, see (6a) above again. On the other hand, it has to capture the fact that the binding of the reciprocal is legitimate within a possessive DP even in the presence of the definite article when the nominal head is a complex event nominal.

---

[9] Recall that a reflexive pronoun is felicitous within the very same environment, which is due to the fact that this pronoun is used logophorically here, cf. (6a) and (i).
(i) *Jól ismerjük* [$_{\text{DP}}$ *a* *magunk* *baj-á-t*].
well know.1PL the ourselves.NOM problem-POSS.3SG-ACC
'We know our own problem well.'

3.1. Encoding the binding domain for reciprocals

As regards the first goal, the crucial aspect of our solution is as follows. We encode the blocking function of the definite article by introducing a special feature: "binding domain delimiter": BDD. We associate this feature with the lexical form of the article in case it occurs in a possessive DP, see (7).

(7) *a(z):* …
(↑CHECK _POSS-MORPH)=c +
(↑BDD)= +

This pair of annotations is optionally assigned to the article, and the XLE-style CHECK feature ensures that the article has this binding domain delimiting function only in possessive DPs. This feature is indispensable for the analysis of Hungarian DPs in general. For instance, it is this feature, encoded by possessive morphology, that licenses the presence of the POSS grammatical function in a DP.[10]

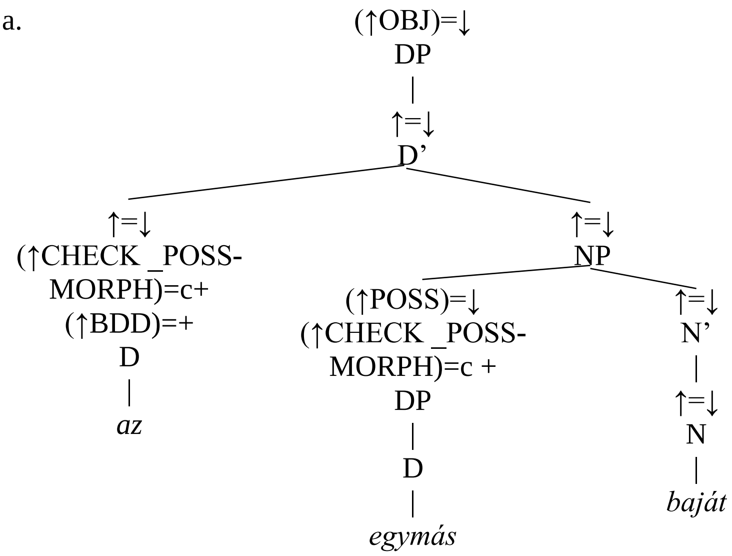
As has been demonstrated in section, Hungarian reciprocals are subject to the Minimal Finite Domain Condition, which must be encoded in their lexical forms. In our analysis this encoding must be coupled with the BDD feature as a negative off-path constraint, see (8). This feature is added as a negative off-path constraint on the domains that involve possessive DPs: the path leading to the anaphor cannot contain this feature. For instance, this renders (6a) ungrammatical in the presence of the article, and the construction is grammatical in the absence of the article.

(8) *egymás:* (GF* GF_{pro})
~(→ TENSE)
~(→ BDD)

In this analysis, the c-structure and f-structure representations of the object possessive DP in (6a) with an overt definite article are as shown in (9a) and (9b), respectively.

[10] The primary function of this particular feature is to check whether the noun head has possessive morphology. For discussions of how XLE-style CHECK features work technically and for their use in the analysis of various Hungarian phenomena, see Laczkó & Rákosi (2011) and Laczkó (2014).

(9) a.



b.
$$\left[\begin{array}{lll} \ldots & & \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{‘problem < (POSS) >’} \\ \text{POSS} & [\text{“each other”}]_i \\ \text{DEF} & + \end{array}\right] & \\ \text{CH\_P-M} & + & \\ \text{BDD} & + & \end{array}\right]$$

3.2. The treatment of reciprocals in possessive event nominals

The analysis as developed so far provides a suitable formal treatment of the facts represented by (6a). However, at this stage its prediction is that the construction type exemplified in (6c) will also be ungrammatical in the presence of the definite article, because the article will have the same blocking effect as in the case (6a), and the binding of the reciprocal by the matrix subject will not be possible, contrary to fact: on the one hand, the construction is grammatical, and, on the other hand, the reciprocal and the matrix subject are coreferential. Our solution, which is the second major aspect of our proposal, is that in the case of this construction type we assume that there is a (covert) local binder for the reciprocal within the possessive DP itself, and this local binder, in turn, is controlled by the matrix subject.

This account capitalizes on Laczkó's (2004) analysis of control relations in Hungarian possessive DPs headed by complex event nominals. First, Laczkó (2004) offers an assessment of the most important previous LFG treatments of GFs in Hungarian possessive DPs: Laczkó (1995), Komlósy (1998), and Chisarik & Payne (2003), and then he argues for an approach in which there are two [–r] function in these DPs: POSS and SUBJ. In this system both these functions can be realized by either the nominative or the dative possessor (which are in complementary distribution). POSS is always overt, and SUBJ is either overt or covert. In the latter case an LFG-style PRO receives this function. Consider Laczkó's (2004:328-331) analysis of the examples in (10)-(12). In the glosses, DEV stands for "deverbal nominalizing suffix".

(10)	a.	*János*	*kiabál-ás-a*
		John.NOM	shout-DEV-POSS.3SG
		'John's shouting'	

	b.	*János-nak*	*a*	*kiabál-ás-a*
		John-DAT	the	shout-DEV-POSS.3SG
		'John's shouting'		

Both the nominative possessor in (10a) and the dative possessor in (10b) are assumed to have the SUBJ function. In (11) the covert agent argument of the nominal is realized by a SUBJ PRO, and Laczkó assumes that it is anaphorically controlled by the matrix subject.[11] Compare (11) with (12), in which the complement of the matrix verb is an infinitival construction.[12]

(11)	*János*	*elkezd-t-e*	*a*	*kiabál-ás-t.*
	John.NOM	start-PAST-3SG.DEF	the	shout-DEV-ACC
	'John started the shouting.'			

(12)	*János*	*elkezdett*	*kiabál-ni.*
	John.NOM	started	shout-INF
	'John started to shout.'		

Notice that in the case of complex event nominals derived from intransitive verbs it would not be necessary to introduce the SUBJ function,

[11] His main argument for the anaphoric control assumption is that the controller can also have an OBL function.

[12] In this construction type the assumption of functional control is the natural choice, given that in Hungarian the controller can only be the matrix SUBJ or OBJ.

in addition to the POSS function. For instance, Laczkó (1995) assumes that in the nominal domain there is a single [–r] function: the "subject-like" POSS. In his analysis of the construction type in (10) the matrix subject controls a POSS PRO. As Laczkó (2004) points out, complications emerge in the case of transitive nominalization. In an "only-POSS" (or an "only-SUBJ") approach the only [–r] function is assigned to that argument of the nominal predicate which is the DP domain counterpart of the object argument of the input verb, see (13). From this it follows that in this scenario there is no "extra" function available for a PRO in a control configuration, compare (14) and (15).

(13)	*a*	*dal*	*elénekl-és-e*	*János*	*által*
	the	song.NOM	sing-DEV-POSS.3SG	John	by
	'the singing of the song by John'				

(14)	*János*	*elkezdte*	*a*	*dal*	*elénekl-és-é-t.*
	John.NOM	started	the	song.NOM	sing-DEV-POSS.3SG-ACC
	'John started the singing of the song.'				

(15)	*János*	*elkezdte*	*elénekel-ni*	*a*	*dal-t.*
	John.NOM	started	sing-INF	the	song-ACC
	'John started to sing the song.'				

By contrast, on a SUBJ & POSS account all analytical details fall into place. The overt possessor constituent, whether in the nominative or in the dative, can be assumed to have the POSS function and the (anaphorically) controlled PRO can naturally get the SUBJ function, see (14), in which the possessor constituent is in the nominative. And the same SUBJ PRO control can be assumed in the case of intransitive nominalization, see (10).[13]

Laczkó (2019) points out that there is independent support for the POSS and SUBJ duality in DPs coming from Russian. Smirnova and Jackendoff (2017) report in a footnote that, in addition to the absolutely productive pattern of expressing the possessor argument as a noun phrase in genitive case, there is a "semiproductive" alternative strategy available that is limited to pronominal arguments, proper names, some kinship terms and some words for professions. Compare their examples in (16)-(18). (16) demonstrates the productive pattern of transitive nominalization. The patient is realized by a genitive constituent, while the agent is expressed as a constituent in

[13] Laczkó's (2004) explanation for why always only one of the two [–r] functions can be overtly realized in Hungarian possessive DPs is that Hungarian possessive DPs obligatorily employ the head-marking strategy, and the inflectional traits of Hungarian nouns are such that they only accommodate a single overt possessormarking. For the details of the LMT mapping of arguments onto these grammatical functions, see Laczkó (2004).

instrumental case. In the semiproductive pattern, by contrast, the patient has the same realization, while the agent is expressed by a prehead argument with possessive morphology, see (17). This is not a pattern generally available to all kinds of possessors, as the contrast between (17) and (18) shows.

(16) *ispolneni-e* *Ravelj-a* *pianist-om*
performance-NOM Ravel-GEN pianist-INST
'the performance of Ravel by the pianist'

(17) *Pet-in-o* *ispolneni-e* *Ravelj-a*
Peter-POSS-NOM performance-NOM Ravel-GEN
'Peter's performance of Ravel'

(18) **pianist-in-o* *ispolneni-e* *Ravelj-a*
pianist-POSS-NOM performance-NOM Ravel-GEN
'the pianist's performance of Ravel'

Smirnova and Jackendoff (2017) leave it to future research to explore how this special pattern can be accommodated in their analysis of argument realization in Russian nominals, which is a special system of overt case assignment to arguments. Laczkó (2019) claims that a GF-based approach of the SUBJ-and-POSS type can naturally accommodate these Russian facts, because for the treatment of the construction type exemplified in (17) the two arguments we need two core GFs. In addition to the standard genitive realization of one of the two central arguments, the other constituent (the external argument) also has possessive morphological marking, as opposed to the standard oblique realization illustrated in (16).

Our analysis of the binding relations in Hungarian DP is cast in the standard LFG theory of anaphora, see Dalrymple (2001). The syntactic constraints on these relations are expressed in terms of f-structural properties. Following Laczkó (2009), we assume the hierarchy of GFs in (19) for the purpose of capturing the relevant anaphoric relations (this is the joint ranking of GFs from the verbal and the nominal domains).

(19) SUBJ > OBJ > OBJ_{θ} > POSS > OBL > ADJUNCT

For instance, the DPs in (20) and (21) are analyzed in our system along the following lines.

(20) *a* *fiú-k* *lefest-és-e* *egymás* *által*
the boy-PL.NOM paint-DEV-POSS.3SG each_other by
'the painting of the boys by each other'

(21)	**egymás*	*lefest-és-e*	*a*	*fiú-k*	*által*
	each_other	paint-DEV-POSS.3SG	the	boy-PL.NOM	by
	'*each other's painting by the boys'				

In both (20) and (21), the two arguments of the nominal are co-arguments, and the reason why (20) is grammatical is that the possessor, which has the SUBJ GF in our system, functionally outranks the OBL argument. By contrast, the (lower-ranked) OBL in (21) cannot bind the reciprocal SUBJ.

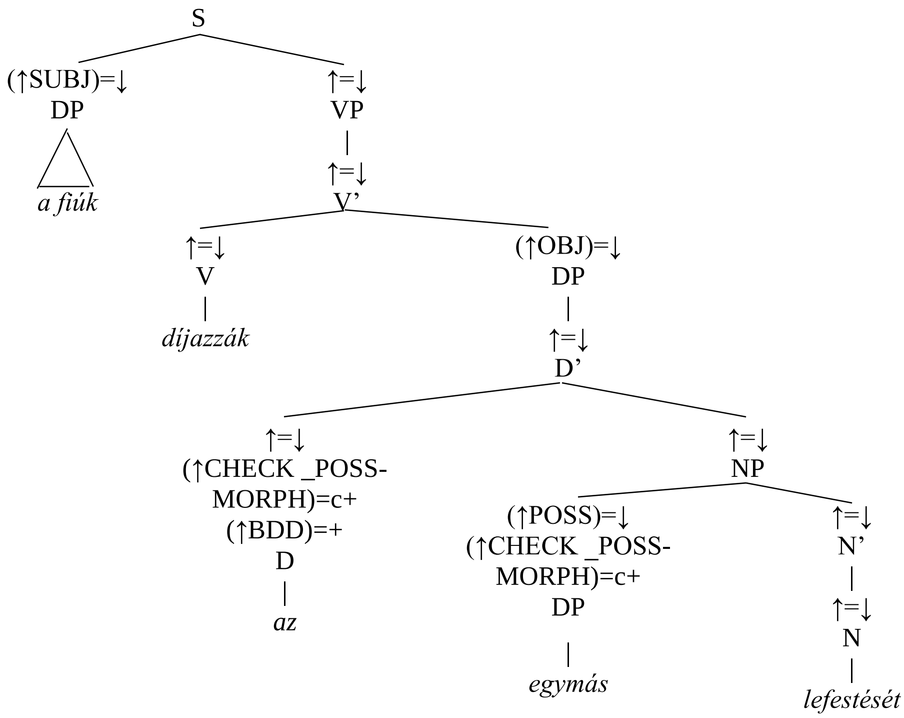
Consider (22) next. Here we assume that the reciprocal anaphor has the POSS function, and it is bound by the higher-ranked SUBJ PRO, which, without any controller in this sentence, has the PROarb interpretation. Notice that without this SUBJ PRO binder the reciprocal could not be treated in an unmarked fashion in LFG's binding theory.

(22)	*Fontos*	(*az*)	*egymás*	*lefest-és-e.*
	important	the	each_other.NOM	paint-DEV-POSS.3SG
	'Painting each other is important.'			

Now let us turn to our crucial example in (6c). In (23) we show our c-structure analysis of the version of this sentence that contains the definite article. In (24) we present the considerably simplified f-structure, where CH_P-M stands for CHECK_POSS-MORPH, and the indices indicate the binding relations.

(6c)	*A fiúk*$_i$	*díjazzák*	[$_{DP}$ *az egymás*$_i$	*lefest-és-é-t*].
	the boys	appreciate.3PL	the each_other	paint-DEV-POSS.3SG-ACC
	'The boys appreciate the painting of each other.'			

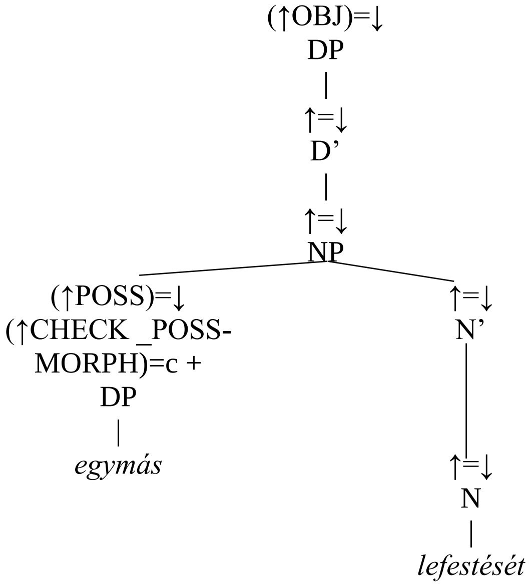
(23)



(24)
$$\left[\begin{array}{ll}
\text{PRED} & \text{‘appreciate < (SUBJ) (OBJ) >’} \\
\text{SUBJ} & [\text{“the boys”}]_i \\
\text{OBJ} & \left[\begin{array}{ll}
\text{PRED} & \text{‘painting < (SUBJ) (POSS) >’} \\
\text{SUBJ} & [\text{“pro”}]_i \\
\text{POSS} & [\text{“each other”}]_i
\end{array}\right] \\
\text{CH _P-M} & + \\
\text{BDD} & +
\end{array}\right]$$

When (6c) does not contain the definite article, the c-structure representation of the possessive DP is as shown in (25).

(25)



The f-structure is the same as in (24), the only difference being that it does not contain the (BDD) feature.

It is important to note that (6c), again, strictly in the presence of the definite article, has another possible interpretation, see (6c'). On this reading the boys appreciate that some other people paint each other.[14] In more technical terms, the antecedent of the reciprocal is different from (i.e. noncoreferential with) the matrix subject. We claim that the crucial aspect of the analysis of this example is the same as that of the analysis of (22): there is a SUBJ PRO antecedent for the reciprocal within the possessive DP.

(6c') *A fiúk*$_i$ *díjazzák* [$_{DP}$ *az egymás*$_k$ *lefest-és-é-t*].
the boys appreciate.3PL the each_other paint-DEV-POSS.3SG-ACC
'The boys appreciate the painting of each other.'

It is also interesting to take a look at an example that illustrates a case when both control and binding are involved, see (26).

(26) *A fiúk*$_i$ *elkezdték* [$_{DP}$ *az egymás*$_i$ *lefest-és-é-t*].
the boys started the each_other paint-DEV-POSS.3SG-ACC
'The boys started the painting of each other.'

[14] See Szűcs (2019) for pertinent discussion.

The f-structure representation of this example is exactly the same as that of (6c) in (24). The only technical difference is that the relationship between the matrix subject and the SUBJ PRO in the case of (6c) is binding, while here it is anaphoric control.

3.3. A note on reflexives

Consider the following example, which is a control construction involving a reflexive in the possessive DP.

(27) *A fiúk*$_i$ *elkezdték* [$_{DP}$ *a* *maguk*$_i$ *lefest-és-é-t*].
the boys started the themselves paint-DEV-POSS.3SG-ACC
'The boys started the painting of themselves.'

Our empirical generalization about Hungarian reflexives above was that, on the one hand, they are subject to Minimal Complete Nucleus Condition, and, on the other hand, they can also be used logophorically. In the case of constructions like (27), it would not at all be appropriate to assume that the coreference between the possessor reflexive and the matrix subject is logophoric in nature, because the covert subject of the derived nominal head is obligatorily controlled by the matrix subject, and it, in turn, obligatorily binds the possessor reflexive. Consequently, if the logophoric analysis is not plausible then the remaining option is the anaphoric treatment. However, in that case the binding domain delimiting function of the definite article, which we assume to hold generally, would block this binding relation. From this it directly follows that even for the treatment of the behaviour of reflexive pronouns in such constructions our approach provides the suitable formal framework: the possessive DP contains a SUBJ PRO, which binds the reflexive, and, in turn, it is controlled by the matrix subject.

4. Summary

In this paper we have dealt with anaphoric pronouns. Partially on the basis of novel data, we have made the following empirical generalizations. The primary reflexive can be used either anaphorically or logophorically, and in its anaphoric use it is subject to the Minimal Complete Nucleus Condition. The reciprocal can only be used anaphorically, and the Minimal Finite Domain Condition applies to it. When either the reflexive or the reciprocal pronoun occurs within a possessive DP, neither of them can be anaphorically bound from outside if this DP contains the definite article, i.e. the article always creates a boundary for the relevant binding domain.

We have developed an LFG analysis of these facts that has two crucial aspects to it. On the one hand, we employ a new feature: BDD ("binding domain delimiter"). We associate this feature with the lexical form of the definite article, and we use it as a negative off-path constraint in modelling the relevant binding relations. On the other hand, following Laczkó (2004, 2009), we assume that within Hungarian possessive DPs there are two [–r] grammatical functions available to arguments of complex event nominals: POSS and SUBJ. Both can be overtly realized by either the nominative or the dative possessor constituent, and, in addition, SUBJ can also be PRO. Thus, we create a DP-internal antecedent for the anaphors in a principled manner, which, in turn, can be controlled from outside the DP. As a result, the binding domain delimiting function of the definite article is still endorsed, and, at the same time, coreference across the article is made possible by the anaphoric control of the SUBJ PRO within the DP.

The postulation of POSS and SUBJ in DPs is necessary for an adequate treatment of control relations, see Laczkó (2004), and it is also necessary for an adequate treatment of binding, see our analysis in this paper. Thus, two phenomena, control and binding, independently and mutually necessitate and support the POSS and SUBJ approach. Furthermore, on the basis of Smirnova & Jackendoff (2017), we have shown that certain data from Russian noun phrases can also be argued to call for the use of both these functions in the nominal domain.

Acknowledgements

Project no. 111918 (*New approaches in the description of the grammar of Hungarian pronominals)* has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K funding scheme.

We gratefully thank our anonymous reviewers for their very helpful and useful comments. All remaining errors are our sole responsibility.

References

Bertocchi, Alessandra & Casadio, Claudia. 1980. Conditions on anaphora: An analysis of reflexives in Latin. In Gualtiero Calboli. ed. *Papers on Grammar I*. Universitá di Bologna. 1-46.

Chisarik, Erika and John Payne. 2003. Modelling possessor constructions in LFG: English and Hungarian. Miriam Butt & Tracy Holloway King. eds. *Nominals: Inside and Out*. Stanford: CSLI Publications. 181-199.

Dalrymple, Mary. 2001. *Lexical Functional Grammar. Syntax and Semantics 34.* New York: Academic Press.

Despić, Miloje. 2011. *Syntax in the absence of determiner phrase*. PhD Dissertation. University of Connecticut.

Despić, Miloje. 2015. Phases, reflexives, and definiteness. *Syntax* 18 (3). 201-234.

É. Kiss, Katalin. 1987. *Configurationality in Hungarian*. Budapest: Akadémiai Kiadó.

Komlósy, András. 1998. *A nomen actionis argumentumainak szintaktikai funkcióiról* [On the syntactic functions of the arguments of the nomen actionis]. Manuscript. Budapest: Institute for Linguistics, Hungarian Academy of Sciences.

Laczkó, Tibor. 1995. *The Syntax of Hungarian Noun Phrases: A Lexical-Functional Approach.* Frankfurt am Main: Peter Lang.

Laczkó, Tibor. 2004. Grammatical functions, LMT, and control in the Hungarian DP revisited. In Miriam Butt & Tracy H. King. eds. *Proceedings of the LFG '04 Conference*. Christchurch: University of Canterbury. 313-333.

Laczkó, Tibor. 2009. On the *-Ás* suffix: Word formation in the syntax? *Acta Linguistica Hungarica* 56 (1). 23-114.

Laczkó, Tibor. 2014. An LFG analysis of verbal modifiers in Hungarian. In: Butt, Miriam & King, Tracy Holloway. eds. *Proceedings of the LFG14 Conference*. Stanford, CA: CSLI Publications, 346-366.

Laczkó, Tibor. 2019. On the case of argument realization in Russian, Hungarian and English nominals. Poster presentation at the *24th International LFG Conference*, at Australian National University, Canberra, July 9, 2019.

Laczkó, Tibor & Rákosi, György. 2011. On particularly predicative particles in Hungarian. In: Butt, Miriam & King, Tracy Holloway. eds. *Proceedings of the LFG11 Conference*. Stanford, CA: CSLI Publications, 299-319.

Marácz, László. 1989. *Asymmetries in Hungarian*. PhD dissertation. University of Groeningen.

Marelj, Marijana. 2011. Bound-variable anaphora and Left Branch Condition. *Syntax* 14 (3). 205-229.

Pollard, Carl & Ivan A. Sag. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry* 23 (2). 261-303.

Rákosi, György. 2010. On snakes and locative binding in Hungarian. In Miriam Butt & Tracy Holloway King. eds. *Proceedings of the LFG '10 Conference.* Stanford: CSLI Publications. 395-415.

Rákosi, György. 2017. The definite article and anaphoric possessors in Hungarian. In *Linguistica Brunensia* 65 (2). 21-33.

Rákosi, György. To appear. With or without the definite article: On the syntax of anaphoric possessor strategies in Hungarian. In Veronika Hegedűs & Irene Vogel (eds.) *Approaches to Hungarian 16*. Amsterdam/Philadelphia: John Benjamins.

Reuland, Eric. 2007. Binding conditions: How can they be derived? Lectures on Binding. Department of Linguistics, St Petersburg University. Joint PhD program St Petersburg–Utrecht University. April 24-May 3, 2007. accessed at http://slioussar.narod.ru/reuland2007.html

Reuland, Eric. 2011. *Anaphora and Language Design*. Linguistic Inquiry Monograph 62. The MIT Press: Cambridge, MA.

Smirnova, Anastasia & Jackendoff, Ray. 2017. Case assignment and argument realization in nominals. *Language* 93(4). 877-910.

Szűcs Péter. 2019. Remarks on binding and control data in Hungarian Complex Event nominals. *Argumentum* 15. 650-664.

Distributive Possessors in Swedish and Norwegian: Binding, Agreement, and Quantification

Helge Lødrup
University of Oslo

Raj Singh
Carleton University

Ida Toivonen
Carleton University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 170–190

http://csli-publications.stanford.edu/LFG/2019

Keywords: distributivity, Scandinavian, quantifiers, Skolemized choice function

Lødrup, Helge, Singh, Raj, & Toivonen, Ida. 2019. Distributive Possessors in Swedish and Norwegian: Binding, Agreement, and Quantification. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 170–190. Stanford, CA: CSLI Publications.

Abstract

The Scandinavian languages employ an unusual device for expressing distance distributivity: they make use of prenominal distributive possessors. These distributive elements appear, at least historically, to be composed of a distributive quantifier and a reflexive possessor. All Scandinavian languages have distributive possessors, but they display some interesting differences across language varieties. Two varieties from Norwegian and Swedish are specifically considered here. We outline similarities and differences between the distributive possessors having to do with agreement, (in-)definiteness, binding, and other linguistically significant properties. We suggest that their interpretive similarities follow from the assumption that they both have the semantics of Skolemized Choice Functions; this assumption makes sense of the fact that they are interpreted as indefinites and as bound variables. We furthermore argue that their main morphosyntactic differences boil down to whether the distributive expression consists of two lexical items or one, following an idea in Vangsnes (2002a,b). Specifically, we propose that the differences follow from the assumption that the Norwegian distributive possessor is a syntactically more complex DP than the Swedish one (the Norwegian variant contains an additional QP that hosts the distributive element).

1 Introduction

The Scandinavian languages have two options for expressing *distance distributivity*: they can use a numeral/indefinite followed by a common noun followed by a distributive element, as illustrated in the Swedish example in (1), or they can use a prenominal distributive possessor (Przepiórkowski & Patejuk 2013; Dotlačil 2012; Zimmermann 2002, a.o.), as illustrated in the Swedish example in (2). Throughout the paper, we mark each Swedish example with an (S) and each Norwegian example with an (N) at the end of the translation line.

(1) Pojkarna har ätit **ett** äpple **var**.
boys.DEF have eaten one apple each
'The boys have eaten one apple each.' (S)

(2) Pojkarna har ätit **varsitt** äpple.
boys.DEF have eaten each.3REFLPOSS.NEUT apple
'The boys have eaten one apple each.' (S)

These constructions are said to express 'distance distributivity' because the sentences are interpreted as something like 'each of the boys has eaten an apple' even though the distributive elements *var* and *varsitt* are far away from the subject *pojkarna* 'boys.DEF'. The strategy in (1) is far more common cross-linguistically (witness the English *a book each*). The focus of our paper is distributive possessors, which are typologically uncommon (but see Dubert & Galves 2016, 422 on Galician).

According to standard criteria for constituency, *varsitt äpple* (containing the distributive possessor *varsitt*) is a syntactic constituent. For example, it can be topicalized as in (3):

(3) Varsin bok har de läst.
each.3REFLPOSS.COM book have they read
'One book each, they have read.' (S)

Examples that include distributive possessors involve matching the so-called *sorting key* (the boys) and the *distributed share* (apple) at a distance, similar to binominal *var* 'each' in (1). This explicit matching differentiates distributive possessor phrases from other similar expressions that also include distributive quantifiers:

(4) Varje pojke har läst en bok.
each boy has read a book
'Each boy has read a book.' (S)

(5) Hon läste varje bok noggrant.
she read each book carefully
'She read every book carefully.' (S)

Sentences with distributive possessors are very similar in meaning to examples such as *Each boy read their book*, where the second NP has a possessor that is bound by the quantified first NP.

Even though distributive possessors are cross-linguistically rare, they occur in all the Scandinavian languages, where they display interesting morphological and syntactic variation. See Faarlund et al. 1997, 207-8, Vangsnes 2002a,b on Norwegian, Teleman et al. 1999, 387-89, Hultman 2003, 120-21 on Swedish, Thráinsson 2001, Sigurðsson et al. Forthcoming on Icelandic, Thráinsson et al. 2004, 129 on Faroese, and Allan et al. 1995 on Danish.

In this paper, we describe and compare distributive possessors in two of the Scandinavian varieties, which we call Standard Swedish and Eastern Norwegian. Norwegian splits into two major dialects that differ in several ways when it comes to distributive possessors (Vangsnes 2002a,b). The Norwegian dialects that we do not discuss are more similar - but not identical - to Swedish in relevant respects. We make sporadic reference to variation that reaches beyond Standard Swedish and Eastern Norwegian, but we wish to stress that this paper is not intended to be a full dialectal survey of distance distributivity in Swedish and Norwegian.

2 The distributive possessor

The distributive possessor consists of a distributive quantifier *hver/var* 'each' and a possessive pronoun. In Norwegian, its first part *hver* is the regular distributive quantifier that translates to *each* in examples such as *hver gutt* 'each boy'. Swedish is different in this respect. In modern Swedish, the regular form for the prenominal distributive quantifier is the uninflected *varje*. Prenominal *var* occurs as well, but only rarely. It can be found dialectally, in older texts, and in certain expressions (e.g. *var sak på sin plats* 'each thing in its place').

In Norwegian, the quantifier *hver* and the possessor are written separately, whereas they can be written separately (*var sin*) or together (*varsin*) in Swedish. We write the Swedish distributive possessor as one word throughout the paper, except we follow the authors in attested examples.

In example (2) above, the sorting key is a subject and the distributed share is an object. However, the NPs can also occur in other positions. The sentences in (6–7) below illustrate the sorting key and the distributed share in various clausal positions.[1] The sorting key phrases are in boldface, and the the distributed share phrases are in small caps. The sorting key phrase is usually a plural NP or pronoun, but it can also be a group-denoting noun, as in (6b):

(6) a. **Elevene** presenterte fakta om HVER SIN
students.DEF presented facts about each.COM 3REFLPOSS.COM
PLANET.
planet
'The students presented facts about one planet each.' (N)

b. **Samboerpar** ble pålagt å ligge i HVERT
cohabitant.couple was instructed to lie in each.NEUT
SITT ROM.
3REFLPOSS.NEUT room
'A cohabitant couple was instructed to have separate rooms.' (N)

c. Du bør gi **dem** HVERT SITT BUR.
you ought.to give them each.NEUT 3REFLPOSS.NEUT cage
'You should give them one cage each.' (N)

(7) a. **Tre lyckliga vinnare** får i veckan nycklarna till
three happy winners get in week.DEF keys.DEF to
VARSIN FORD THUNDERBIRD 1955.
each.3REFLPOSS.COM Ford Thunderbird 1955
'Three lucky winners will this week get the keys to one 1955 Ford Thunderbird each.' (S)

b. Efter denna kanonad stannar matchen av och avslutas med
after this bombardment stops game.DEF off and finishes with
ytterligare VARSITT MÅL för **de båda lagen**.
additional each.3REFLPOSS.NEUT goal for the both teams.DEF
'After this bombardment, the game finishes with one additional goal each for both teams.' (S)

[1]Almost all of the examples in this paper are attested examples retrieved from the world wide web, either directly with Google, or indirectly through corpora of web texts. Some examples have been shortened or modified slightly, but not in a way that is relevant to the points we make.

c. Jag gav **dom** VARSIN MOROT.
I gave them each.3REFLPOSS.COM carrot
'I gave them one carrot each.' (S)

The examples above show that neither the sorting key nor the distributed share is tied to a particular phrase structural position or grammatical function. The distribution is not unrestricted, however; we will return to this in Section 2.5 below.

2.1 Agreement

The distributive possessor displays richer agreement in Eastern Norwegian than in Swedish. In Eastern Norwegian, both the quantifier and the possessor agree, but only the possessor agrees in Swedish.

In Eastern Norwegian, the quantifier agrees with the distributed share. The possessor agrees both with the sorting key and with the distributed share. These facts are illustrated in 8–9 and discussed immediately below:

(8) Guttene fikk **hver** **sin** sykkel.
boys.DEF got each.COM 3REFLPOSS.COM bike(COM)
'The boys got one bike each.' (N)

(9) Vi fikk **hvert** **vårt** bord.
we got each.NEUT our.NEUT table(NEUT)
'We got one table each.' (N)

The possessor agrees with the sorting key in person and number, a case of index agreement (Wechsler & Zlatić, 2000). In (8), it agrees with the third person *guttene* 'boys.DEF', and in (9) with the first person plural *vi*. Note that the Scandinavian languages have separate reflexive forms in the third person only; the first and second person forms are used both reflexively and non-reflexively.

The possessor also agrees with the distributed share in gender and number, a case of concord agreement (Wechsler & Zlatić, 2000). It agrees with the common gender noun *sykkel* 'bike' in (8) and with the neuter *bord* 'table' in (9). The quantifier also agrees with *sykkel* and *bord* in gender. When the distributed share is plural, as in (21) below, the morphologically unmarked common gender form is used.

Compare Eastern Norwegian (8–9) to the parallel Swedish in (10–11):

(10) Pojkarna fick **varsin** cykel.
boys.DEF got each.3REFLPOSS.COM bike(COM)
'The boys got one bike each.' (S)

(11) Vi fick **varsitt** bord.
we got each.3REFLPOSS.NEUT table(NEUT)
'We got one table each.' (S)

In (10–11), the Swedish possessor agrees with *cykel* 'bike' and *bord* 'table', but the quantifier *var* does not.[2]

Outside the distributive possessor construction, both Eastern Norwegian and Swedish possessive pronouns show index agreement with the possessor. Similarly, the prenominal quantifier *hver/var* agrees with the noun it quantifies outside the distributive possessor construction (in Swedish this is the case only when the quantifier is *var*; the distributive quantifier is usually the non-inflecting *varje*, as in (4)–(5) above).

2.2 Definiteness

Possessive NPs are in general definite (Lyons 1999, 1.2.4, Barker 2000; Peters & Westerståhl 2013), but distributive possessor phrases seem to not be: they can occur in contexts normally restricted to indefinites. One example is the object position in presentational sentences, as in Eastern Norwegian (12) and Swedish (13):

(12) Det ble overrakt dem **hver** **sin** **medalje**.
it was given them each.COM 3REFLPOSS.COM medal
'They were given one medal each.' (N)

(13) Det ligger **var** **sin** **skattkarta** till barnen redo hemma.
there lies each 3REFLPOSS.COM treasure.map to children.DEF ready home
'There is one treasure map each for the children at home.' (S)

The distributive possessor phrases cannot felicitously be exchanged for possessive or other definite NPs in (12–13). The indefinite nature of Scandinavian distributive possessor phrases is unsurprising in light of the fact that distance distributivity marking cross-linguistically appears on indefinite NPs (Safir & Stowell, 1988; see also Milačić et al., 2015 as well as Section 3.2 below for an attempt to explain this generalization).

In Swedish, some dialects allow the indefinite article, homophonous with the numeral 'one', to precede the distributive possessor: *en varsin* and *ett varsitt*.

(14) Alla barn får ett paket med **en** **varsin** bok i.
all children get a package with one each.3REFLPOSS.COM book in
'All children receive a package with one book each in it.' (S)

[2]Examples of quantifier agreement in Swedish occur but are infrequent (i). Hultman (2003, 120) refers to quantifier agreement in distributive possessors as hypercorrection.

(i) Sedan gav brudgummen oss vartsitt kuvert med pengar i!
then gave groom.DEF us each.NEUT.3REFLPOSS.NEUT envelope with money in
'Then the groom gave us one envelope each with money!' (S)

Regardless of whether *en/ett* is interpreted as an indefinite article or the number 'one', *en/ett* phrases are indefinite. Examples similar to (14) but with the definite article *den/det* instead of *en/ett* do not occur: **den varsin bok*.

Other Swedish dialects have reanalyzed *varsin* and *varsitt* as *vars* plus the indefinite article (or the numeral 'one'): *vars en* and *vars ett*. An example is (15):

(15) Alla elever ska ha **vars** **en** **bok**.
all pupils shall have each.GEN one book
'All pupils must have one book each.' (S)

These dialects, which seem to be spoken mainly in Scania, also have *vars två* 'each.POSS two', *vars tre* 'each.POSS three', etc.:

(16) Till sist gick vi till Andrahandsbokhandeln där vi båda hittade **vars** **två** **böcker**.
to last walked we to Second.hand.bookstore.DEF there we both found each.GEN two books
'Finally, we went to the second hand book store, where we both found two books each.' (S)

The fact that the dialectal forms include *en/ett* and (other) numerals further indicates that the distributed share is indefinite in Swedish.

2.3 Attributive adjectives

Both Norwegian and Swedish have a distinction between what is traditionally called 'weak' and 'strong' adjective declension. The former is used in definite NPs, and the latter in indefinite NPs, as shown in Norwegian (17) and Swedish (18). This is typically considered to be definiteness agreement. We will use the terms definite and indefinite about these adjective forms. Definite adjectives do not agree in number or gender.

(17) a. den **lange** boken
the.COM long.DEF book.DEF
'the long book' (N)

b. min /sin **lange** bok
my.COM /3REFLPOSS.COM long.DEF book
'my long book' (N)

c. en **lang** bok
a.COM long.INDEF.COM book
'a long book' (N)

(18) a. den **långa** boken
the.COM long.DEF book.DEF
'the long book' (S)

b. min /sin **långa** bok
my.COM /3REFLPOSS.COM long.DEF book
'my long book' (S)

c. en **lång** bok
a.COM long.INDEF.COM book
'a long book' (S)

In Norwegian, adjectives display definite forms in distributive possessor phrases (19), but in Swedish, adjectives display indefinite forms in distributive possessor phrases (20):[3]

(19) a. Vi leste hver vår **lange** bok.
we read each.COM our.COM long.DEF book
'We read one long book each.' (N)

b. Elevene lager hver sin **lille** skulptur.
studentsDEF make each.COM 3REFLPOSS.COM little.DEF sculpture
'The students made one little sculpture each.' (N)

(20) a. Vi läste varsin **lång** bok.
we read.PAST each.3REFLPOSS.COM long.INDEF.COM book
'We read one long book each'. (S)

b. De bär på varsin **liten** sändare
they carry on each.3REFLPOSS.COM little.INDEF.COM transmitter
och mottagare.
and receiver
'They carry one little transmitter and receiver each.' (S)

We will argue in Section 3.2 that *hver* is the source of indefiniteness in Norwegian distributive possessor phrases. Furthermore, we will argue in Section 3.1 that *hver* is the specifier of the DP containing the adjective and *vår* is the D head: $[_{DP}$ $[_{QP}$ *hver*$]$ $[_{D'}$ *vår lange bok*$]]$. Hence the adjective can be definite (because possessives are definite) even though the entire distributive possessive phrase is indefinite (once the contribution of "hver" is factored in). In Swedish *varsin* is the head of the DP containing the adjective, $[_{DP}$ *varsin lång bok*$]$, and is also the source of indefiniteness (again, see Section 3.2). Hence, the distinction between Swedish and

---

[3]None of the Swedish speakers we have consulted accept definite adjectives in distributive possessor phrases. However, some attested examples do occur: the example below is from Ivar Lo-Johansson's (1985) *Frihet*, the fourth volume of his memoirs. Definite forms after *varsin* in Swedish are unusual, and we do not know what governs the variation.

(1) Vi drack varsin lilla kopp espresso.
we drank each.3REFLPOSS.COM little.DEF cup espresso
'We drank one small cup of espresso each.' (S)

Norwegian declensions follows from whether the source of indefiniteness heads the DP that hosts the adjective.

2.4 Number

Prototypically, a sentence with a distributive possessor has a singular distributed share. Teleman et al. (1999, 388) say that the distributed share is “normally" singular in Swedish. Searches in the Norwegian web-corpus NoWaC show that plural *hver sine* make up only 6.6% of the total number of ‘hver lemma’ + ‘sin lemma’. Some speakers allow a plural noun preceded by a numeral as the distributed share. In this case, the numeral decides *how many Xs each* are intended.

(21) De har tenkt å male hver sine to rom
they have thought to paint each 3REFLPOSS.PL two rooms
‘They intend to paint two rooms each.’ (N)

(22) Efter det kommer Italien och Tyskland med varsina fyra vinster.
after it come Italy and Germany with each.3REFLPOSS.PL four wins
‘After that come Italy and Germany with four wins each.’ (S)

However, many speakers do not accept such sentences. In a query, the Norwegian (21) got an average acceptability score of about 2.5 out of 5, and its Swedish counterpart about 2.1. By comparison, all examples included in the survey that contained the singular *hver sin/varsin* received scores higher than 4.8.

Both Swedish and Norwegian allow sentences with the *one X each* interpretation to have a plural distributed share, as an alternative to a singular distributed share. Consider the Swedish example (23):

(23) Pojkarna gick till varsina rum.
boys.DEF went to each.3REFLPOSS.PL rooms
‘The boys went to one room each.’ (S)

In example (23), the plural could be replaced by the singular without any change in interpretation. Even if a plural is possible, it is clear that the singular is the unmarked choice. In the query mentioned above, (23) had an average acceptability score of about 2.5 out of 5, and the corresponding Norwegian sentence about 3.

There is an interesting difference in interpretation between the Norwegian and the Swedish plural distributed share. In Swedish, the natural interpretation of *varsina X-pl* is ‘one X each’, even though the phrase is in the plural. In fact, the Swedish speakers we have consulted seem to think this is the only possible interpretation.[4] However, Norwegian speakers find the phrase ambiguous between ‘one

[4] There does, however, seem to be variation in Swedish concerning the interpretation of *varsina X-pl*. Hultman (2003, 120) gives the impression that Swedish is like Norwegian.

X each' and 'some Xs each'. There is no preference for the singular; if anything, the plural interpretation is preferred, unless context dictates otherwise.

2.5 Binding

A striking difference between Eastern Norwegian and Swedish concerns binding. Eastern Norwegian *hver sin* must be syntactically bound in the same way as the regular reflexive possessive *sin* (see Faarlund et al. 1997, 1154, Vangsnes 2002a), while this is not necessarily the case in Swedish. The question is then how the regular reflexive possessive is bound. Binding conditions are basically the same in the Mainland Scandinavian languages (but see Lundquist 2014 for some nuances). The regular reflexive possessive *sin* is the possessive of both the simple reflexive *seg/sig* and the complex reflexive *seg selv/sig själv*. The distribution of the regular reflexive possessive *sin* is therefore the union of the distribution of the simple and the complex reflexive. Its binding domain is the minimal finite domain (Hellan 1988, 59-79, Dalrymple 1993, 32-33). Somewhat more controversial is the question of possible binders. In our view, the Scandinavian literature on binding tends to be too restrictive concerning the options that actually exist. Without going into details, we would like to point out that not only subjects, but also objects are to some extent possible binders of regular reflexives (see e.g., Platzack 1998, 222-23 on Swedish, Lødrup 2008 on Norwegian); an example is (24). It is also true of Swedish and Norwegian *varsin* and *hver sin*; see examples (7c) and (6c) above.

(24) Jeg ga dem maten sin.
I gave them food.DEF 3REFLPOSS.COM
'I gave them their food.' (N)

In Eastern Norwegian, *hver sin* has the same binding domain as the regular reflexive possessive *sin*, namely the finite domain. Swedish *varsin* differs from its Eastern Norwegian counterpart in that it sometimes allows a binder that does not satisfy regular binding conditions. In some cases, the binder does not outrank *varsin*, as in (7b) and (13) above, and (25) and (26) (the latter from Teleman et al. 1999, 388). In other cases, the binder is not syntactically realized, as in (27).[5]

(25) **Var sin kopp kaffe** lockade två lyssnare till Hasses
each 3REFLPOSS.COM cup coffee tempted two listeners to Hasse's
lilla trädgårdshörna.
little garden.corner
'One cup of coffee each tempted two listeners to Hasse's little garden corner.' (S)

[5]By 'syntactically realized' we here mean *overtly* syntactically realized. The word *frukost* 'breakfast' in (27) presumably introduces (implicit) event participants, and the sentence asserts that there was a cup of coffee for each of them. Clearly, more would need to be said about what notion of "syntactic realization" this is, and how it relates to the binding theory in general.

(26) Jag gav **var sin** **båt** åt mina bröder.
I gave each 3REFLPOSS.COM boat to my brothers
'I gave my brothers one boat each.' (S)

(27) Till frukost idag blev det smörgås, och så klart
to breakfast today became it sandwich and of course
varsin **kopp kaffe**.
each.3REFLPOSS.COM cup coffee
'For breakfast today, there were sandwiches, and of course one cup of coffee each.' (S)

It is noteworthy that (26) is given as a regular example in the reference grammar of the Swedish Academy (Teleman et al. 1999, 388). The Norwegian variants of the Swedish sentences that do not satisfy standard binding requirements give the impression of being degraded. (Scattered Norwegian examples that violate binding conditions can be found in texts, but corpus searches indicate a real difference between Norwegian and Swedish.[6])

2.6 Diachrony

The diachrony of the distributive possessor construction has not been investigated. However, it seems plausible that its origin is sentences with 'each' as a floating quantifier (Faarlund et al. 1997, 207, Askedal et al. 2013, 102-3). Consider the Norwegian (28) - (29). Example (28) has *hver* as a floating quantifier, while example (29) is structurally ambiguous between an analysis with *hver* as a floating quantifier and *hver* as a part of a distributive possessor.

(28) Sjåførene har hver fått sin rute.
drivers.DEF have each.COM got 3REFLPOSS.COM route
'Each driver has got his/her route.' (N)

(29) Sjåførene har hver sin rute.
drivers.DEF have each.COM 3REFLPOSS.COM route
'Each driver has his/her route.' (N)

Sentences in which the floating quantifier precedes the object have likely been reanalyzed to yield the complex distributive possessor. Originally, the floating quantifier must have been the binder of the reflexive possessive, and triggered its agreement in person. The quantifier *hver* 'each' can only be third person. This means that the invariable use of third person *sin* as in Modern Swedish reflects the original situation while Eastern Norwegian person agreement with the sorting key

[6]We searched for *hver sin* in the Norwegian web-corpus NoWaC. The first 200 hits contained one sentence with a clear violation of standard Scandinavan binding conditions. We then searched for *varsin* in the social media corpora at the Swedish Korp corpus collection. The first 200 hits contained twelve sentences with clear violations of binding conditions.

is an innovation. Other Norwegian dialects only use *sin*, see Vangsnes (2002a,b), and this is also the main rule in Danish, see Allan et al. (1995, §5.3.1.8). (Old Norse used the third person reflexive in related constructions with 'each', see Faarlund 2004, 283-84.) When the floating quantifier was the binder, it could not agree with the distributed share. Again, Eastern Norwegian shows an innovation, letting 'each' agree with the distributed share. Other Norwegian dialects only use the morphologically unmarked form (Vangsnes 2002a,b).

3 Analysis

3.1 Syntax

There are a number of reasons to consider Swedish *varsin* as one word and Eastern Norwegian *hver sin* as two. As mentioned above, Swedish *varsin* is often written as one word, while Eastern Norwegian *hver sin* is not. In addition, *varsin* lacks the direct connection to the regular distributive quantifier that Norwegian *hver sin* has: the regular prenominal quantifier is *varje*, not *var*, in Swedish. Dialectal pronunciations such as 'vassin' also indicate the word status of Swedish *varsin*.

The second part of Eastern Norwegian *hver sin* shares important properties with the reflexive possessive: (1) It agrees with the sorting key in person and number. (2) It triggers the definite form of the adjective. (3) It follows binding theory.

The *sin* in Swedish *varsin* lacks these properties. According to Teleman et al. (1999, 387), Swedish *varsin* is lexicalized. Vangsnes (2002b) similarly argues that *hver* and *sin* constitute one "lexical combination" in Norwegian dialects that show the Swedish agreement pattern, while Eastern Norwegian has the quantifier and the possessive as two distinct lexical items.

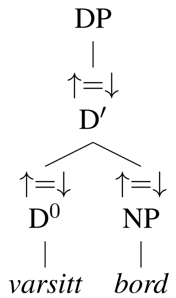
The assumption that Swedish *varsin* is lexicalized explains why 'sin' does not agree with the sorting key and why 'var' does not agree with the distributed share. This assumption also sheds light on the reanalyses in Swedish dialects that have reanalyzed *varsin* as shown in examples (14) and (15) above (with *en varsin* and *vars en* respectively) - this kind of reanalysis seems to presuppose a lexical unit as its point of departure.

Based on the morphological, syntactic, and semantic characteristics presented above, we assume that the Swedish distributive possessor is a single lexical unit where *sin* no longer functions as a possessor. In Eastern Norwegian, the quantifier and the possessor are two separate lexical items. Our analysis is inspired by Vangsnes's (2002a,b) analysis of distributive possessors in Norwegian dialects.

We follow a suggestion by Vangsnes (2002b) that the words in the separable *hvert sitt* occupy a QP in specifier position and a D^o head respectively, whereas the single lexical unit occupies a head position. We further draw upon the syntax for *every* proposed in Dalrymple (2001, section 8.2) and the analysis of pronominal possessors in Strunk (2004). Following Dipper (2005) and Spector (2008), we assume that quantifiers can be of different categories — Spector (2008) specifically proposes that they can be of the category D or Q. The lexical entries and c-structure that we assume are given in (30–31). The subscript DD (for *Distance Distributivity*) in (31a) marks the entry for *hver* 'each' that is used together with a reflexive pronoun and associated with the relevant semantics.

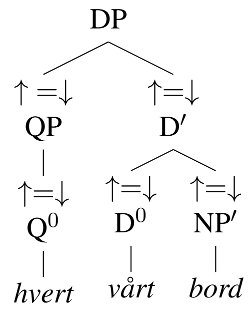
(30) SWEDISH

varsitt	D	(↑ SPEC PRED)	=	'EACH'
		(↑ DISTRIBUTIVE)	=	+
		(↑ CONCORD NUMBER)	=	SG
		(↑ CONCORD GENDER)	=	NEUTER
		(↑ DEFINITE)	=	−



(31) EASTERN NORWEGIAN

a.	*hvert*$_{DD}$	Q	(↑ SPEC PRED)	=	'EACH'
			(↑ DISTRIBUTIVE)	=	+
			(↑ CONCORD NUMBER)	=	SG
			(↑ CONCORD GENDER)	=	NEUTER
			(↑ DEFINITE)	=	−
			(↑ POSS REFL)	=	+
b.	*vårt*	D	(↑ POSS PRED)	=	'pro'
			(↑ POSS INDEX NUMBER)	=	PL
			(↑ POSS INDEX PERSON)	=	1
			(↑ CONCORD NUMBER)	=	SG
			(↑ CONCORD GENDER)	=	NEUTER
			(↑ CONCORD DEFINITE)	=	+



The Eastern Norwegian lexical entries contain more agreement information than the Swedish entry, reflecting the richer agreement marking in Eastern Norwegian. Distributive possessor phrases are indefinite in both Eastern Norwegian and Swedish. The Eastern Norwegian distributive possessor also contributes definite concord agreement, since it includes a possessive pronoun, and Norwegian possessive pronouns require definite concord on adjectives.

In Swedish, the semantics associated with distance distributivity is associated with a specialized lexical entry for *varsitt* (or *varsin*). Eastern Norwegian is more straightforward, since it makes use of words that exist independently – *hver/hvert* and possessive pronouns. However, the specific interpretation associated with distance distributivity only occurs when the *hver/hvert* co-occurs with a reflexive pronoun that is co-indexed with the sorting key. In our semantic proposal in Section 3.2, the distributed share is interpreted as a so-called 'Skolemized Choice Function'. This function is associated with *hver/hvert* precisely when it co-occurs with a possessive pronoun. We capture this formally by assuming that the relevant use of *hver/hvert* (31a) includes the specification (↑ POSS REFL) $= +$. This equation adds a POSS grammatical function which needs to receive its PRED feature from some other lexical entry. The possessor is reflexive, which only allows *sin/sitt* in third person, assuming that other third person pronouns are marked (↑ REFL) $=$ $-$. Our analysis adds a [REFLEXIVE $+$] feature also to first and second person possessive pronouns which do not have morphologically distinct reflexive forms.

3.2 Toward a semantic analysis

Our discussion above imposes several demands on a semantic analysis of distributive possessor phrases like *varsin bok* and *hver sin bok*:

(32) a. They are indefinite noun phrases (even though there do not appear to be any $\exists$-denoting elements in them).

b. They behave anaphorically (we need to find antecedents in order to interpret them).

c. The antecedent is typically a universal quantifier denoted by a plural definite noun phrase (giving rise to a $\forall > \exists$ scope configuration).

We suggest that these demands can be naturally satisfied if we follow Milačić et al. (2015) in assuming that markers of distance distributivity denote Skolemized Choice Functions (SCFs). Roughly speaking, SCFs provide a formal means to 'pair' elements from one set with elements from another. For example, consider a sentence like *the boys ate an apple each*. Intuitively, the sentence is true if there is a way to pair each boy with an apple such that each boy ate the apple he is paired with. Milačić et al. (2015) argued that these truth-conditions (among other relevant facts) could be derived if *an apple each* denotes a pairing function of this kind. We propose that the requirements in (32) can be captured if we assume that distributive possessor phrases – like related phrases like the English *an apple each* or the Swedish *ett äpple var* – also denote SCFs.

Choice functions – Skolemized or not (see below for the distinction) – have been argued to play an essential role in the syntax and semantics of natural language indefinite noun phrases (Reinhart, 1997; Winter, 1997; Matthewson, 1999; Schlenker, 2006; Steedman, 2011, a.o.). We do not discuss their motivation here. What is important for our purposes is that SCFs simultaneously provide variables – and hence the possibility for anaphora (cf. (32b)) – as well as existential meanings without the use of existential quantifiers (cf. (32a)). Together, these formal tools combine in a phrase like *varsin bok* to take each element from some domain (usually given by a plural definite subject) and pair it with a book (cf. (32c)).

We now say a bit more about Choice Functions, just enough to illustrate their application to the semantics of distributive possessives. A *Choice Function* f is a way of picking elements from a set: given a non-empty set P, $f(P) \in P$. Letting f be a choice function variable, a formula like $read(Sara, f(book))$ is true if there is a way f of choosing from the set of books, $f(book)$, such that Sara read $f(book)$.[7] This is just a roundabout way of saying that the sentence is true if there is a book that Sara read. A *Skolemized Choice Function* is a way of mapping individuals to choices from a set. Specifically, given a sequence of individuals $d_1, \ldots, d_k$, and a non-empty set P, a SCF maps these inputs to an element of P: $f(d_1, \ldots, d_k, P) \in P$. Here we say the *arity* of the SCF is k (and hence a 'pure' Choice Function is a nullary SFC).

With this as background, consider a sentence like the Swedish *flickorna läste varsin bok* ('the girls read *varsin* book', i.e., 'the girls read a book each'). Recall from (32) that we need a semantic analysis under which (i) the sentence receives a $\forall\exists$ interpretation (i.e., a quantifier alternation with a universal quantifier outscoping an existential quantifier) and (ii) *varsin bok* behaves like a variable bound by the higher universal quantifier. The classic first-order logic representation for capturing

[7] It is sometimes assumed that choice function variables are closed off by an existential closure operator at matrix level, such that the above formula would actually be a sentence $\exists f(read(Sara, f(book)))$. Note that by the definition of satisfiability, there is 'implicit' existential quantification over the choice function variable when the $\exists f$ is left off. Partly for this reason, and partly to reduce clutter, we omit the existential closure in our representations. Note also that *book* in our representation is really the characteristic set of the function $\lambda x.book(x)$, and so on for other aspects of our representation. We hope our abuse of notation does not lead to any material difficulties.

(i) does not readily give us the means to also capture (ii):

(33) $[\forall x : girl(x)][\exists y : book(y)](read(x, y))$

This representation does not explicitly indicate that the choice of book y depends on the choice of girl x. Furthermore, in thinking about the semantics of phrases like *varsin bok*, it is not clear which element(s) in this phrase might plausibly denote $\exists y : book(y)$ (or a generalized quantifier variant like λP. **book** $\cap P \neq \emptyset$, where **book** is the set of books).

In contrast to this, consider an alternative representation like in (34):

(34) $[\forall x : girl(x)](read(x, f(x, book)))$

Here, unlike (33), there is no explicit existential quantifier (cf. (32a)). Instead, there is a SCF in the second argument of *read* that pairs each girl x with a book $f(x, book)$ (we can think of this, roughly, as *x's book*; cf. (32b) and (32c)).We can think of (34) as a 'Skolemized' variant of (33): one is satisfiable if and only if the other one is, but Skolemization eliminates explicit existential quantification and replaces it with the use of SCFs that explicitly mark the dependence of choices of books on choices of girls.[8] In this way, the representation in (34) acts like a function that 'pairs' elements of one set (the girls) with elements of another set (the books). The formula in (34) is true if there is a unary SCF f that pairs girls and books such that each girl x is paired by f with a book $f(x, book)$ that x read. For now, we assume no constraints on the function f. However, Milačić et al. (2015) suggest that the most natural reading is that it is a one-to-one function, and they note that Swedish arguably requires that it be one-to-one (Teleman et al., 1999).

Our proposal is that Scandinavian distributive possessor phrases *hver sin N* and *varsin N* denote SCFs that pair each element of an antecedent set with some element in the set denoted by *N*. That is:

(35) The denotation of *hver sin N* and of *varsin N* is $f(x, N)$.

With this assumption, let's see how the demands in (32) above can be met and how a representation like (34) can be compositionally derived for a sentence like the Swedish *flickorna läste varsin bok*. First, the existential quantificational force of distributive possessors comes from the assumption that these phrases introduce a SCF variable f which, recall, are undersood as existentials (either implicitly or explicitly – see Note 7). Second, there is a variable x inside the Skolem term $f(x, book)$. We assume that this variable is lexically specified as a bound variable (with possibly different binding domains in different languages). Furthermore, this variable is bound by the higher universal quantifier. For now, we have no way to *derive* the fact that the binder must be a higher universal quantifier that furthermore

[8]Incidentally, one of the main motivations for Skolemization in the mathematical logic literature was to eliminate existential quantifiers (and hence reduce quantifier alternations) while retaining relevant semantic notions like satisfiability. See Buss (1998) for discussion.

is canonically given by a plural definite. This might be the place for a further lexical stipulation that x somehow associates with a covert distributive operator that turns plural definites into universal quantifiers (Note 9). We leave this matter unresolved for now and hope to return to it in future work (see also Section 4).

Consider now the compositional derivation of (34) for *flickorna läste varsin bok* ('the girls read varsin book'). The plural definite subject *flickorna* denotes a universal generalized quantifier over girls: $\lambda P_{et}.[\forall x : girl(x)]P(x)$.[9] The object *varsin bok* denotes a SCF $f(x, book)$; the fact that the Skolem variable x ends up being bound by the higher universal quantifier follows from the lexical assumption associated with *varsin* that its variable is an anaphoric element. Our discussion in Section 2.5 would suggest that the binding domain is the finite domain in Eastern Norwegian, and at least for the simplest cases in Swedish. Thus, assuming a standard lexical entry for *läste* (e.g., $\lambda y.\lambda z.read(z, y)$), the entire sentence composes to yield $[\forall x : girl(x)](read(x, f(x, book)))$, as desired (= (34) above).[10]

4 Discussion

We have presented a series of generalizations about the syntax/semantics of distributive possessors in Eastern Norwegian and Swedish. We have proposed that the syntax of these phrases is different in the two languages, and that Skolemized Choice Functions play an essential role in their interpretation. Here we end by highlighting some challenges that remain for our proposal.

It follows from our semantic analysis that the distributed share needs a sorting key. We noted earlier that the *canonical* sorting key tends to be the denotation of a plural definite noun phrase, but there are other possibilities (see examples (6b) and (7a) above, as well as Milačić et al., 2015). There are two challenges here: what makes plural definites canonical, and which non-canonical sorting keys are

[9]In Milačić et al. (2015), we assumed with Heim et al. (1991) that a covert distributive operator D could apply to the Link (1983) style referential output of plural definites (the maximal object). In place of $\forall x : girl(x)$, then, we assumed the meaning was $\forall x \sqsubseteq MAX(Girl)$, where $\sqsubseteq$ is 'atomic-part-of' and $MAX(Girl)$ is the maximal element in the set $Girl$ ordered by the 'part-of' relation (following Link). One could alternatively assume an ambiguity in the definite itself: either it is referential, or it is a universal quantifier (something would need to be said about presuppositions). Another option is to leave the definite as a purely referential element, and pack the distributivity into the choice functional element instead. For example, a lexical entry like the following – along with suitably type-shifted variants for occurrences in non-canonical positions – would go quite far in capturing the data: $[[varsin]] = \lambda P_{et}.\lambda R_{<e,et>}.\lambda X_e.\exists f \forall x \sqsubseteq X(R(x, f(x, P)))$. This move is undesirable in part because plural definites can receive a distributive quantificational interpretation without *varsin*. There are other choice points as well, such as the actual type of the quantifier. In some treatments, quantificational noun phrases do not denote generalized quantifiers but instead take a variable and two open formulas as input (e.g., Heim, 1982; Dalrymple, 2001; see Heim, 1997 for relevant discussion). Here we simply note that there are many viable approaches for turning a plural definite into a universal quantifier, and we do not commit to any particular way of doing it.

[10]The verb and object combine to give $\lambda z.read(z, f(x, book))$, and these combine with the subject universal quantifier $\lambda P_{et}.[\forall x : girl(x)]P(x)$ to give the final result $[\forall x : girl(x)](read(x, f(x, book)))$.

allowed? Furthermore, there is cross-linguistic variation in exactly what a possible sorting key is. In Eastern Norwegian, for example, the relationship between the distributed share and the sorting key mirrors that of anaphors and their antecedent: the sorting key must outrank the distributed share. Specifically, the binding domain of *hver sin* appears to be identical to that of *sin*. The constraints on the Swedish sorting key are less strict, and the sorting key does not necessarily outrank the distributed share. The Swedish example in (36) is ambiguous: the gifts can distribute over the guest or the children. In other words, if there are three guests and four children, either three or four gifts were given. In the Eastern Norwegian equivalent, the gifts distribute over the guests, not the children.

(36) Gästerna gav **varsitt** **paket** åt barnen.
guest.DEF gave each.3REFLPOSS.NEUT gift to children.DEF
'The guests gave the children one gift each.' (S)

Thus, as noted earlier, one might say that the binding domain of Swedish *varsin* is the finite domain in the simplest cases. However, in Swedish the sorting key is not necessarily included in the sentence at all; it can also be a referent retrieved from the context, as in (27) above. Here, just like when the sorting key is explicit, the implied sorting key is understood as a universal quantifier (for each person at breakfast today x), and the Skolem term pairs each such x with a cup of coffee ($f(x, C)$, where C is a (salient) set of coffee cups). Example (27) is ungrammatical if *varsin kopp kaffe* is replaced by a noun phrase with a regular reflexive possessor. Similarly, if *varsitt paket* in (36) is replaced by a noun phrase with a reflexive possessor (e.g., *sitt paket* 'REFL.POSS gift'), it is unambiguously bound by the subject in both Swedish and Eastern Norwegian.

We hope to put these facts into order and to relate them to our general assumption that *hver sin* is composed of two units while *varsin* is a single unit, as well as the assumption that Skolem term variables might be subject to different anaphoric constraints across languages. Having made assumptions about both indefinites and anaphora, we hope to connect the ideas sketched here to related work on the connection between indefinites and pronouns (e.g., Kamp, 1981; Heim, 1982; Groenendijk & Stokhof, 1991; Crouch & van Genabith, 1999; Dalrymple, 2001), and to embed choice-functional treatments of indefinites within glue-theoretic approaches to semantics (e.g., Dalrymple, 2001).

As but one step in this direction, consider the following proposal for the meaning constructor for Norwegian *hver* as it occurs in *hver sin bok*. Based on our earlier discussion, *hver* might plausibly denote $\lambda x_e.\lambda P_{et}.f(x, P)$ (the fact that the Skolem variable ends up bound would follow from the assumption that *sin* is co-indexed with the subject universal quantifier in the syntax, and that an anaphor that is co-indexed with a quantificational noun phrase that outranks it is interpreted as a variable bound by the quantifier). The glue side would seek a resource of the type supplied by *sin*, and would return an implication that seeks a resource of the type supplied by *bok* to return a resource y_σ (where y is the label for the F-structure for the entire object DP *hver sin bok*). The rest of the composition would follow from standard assumptions (e.g., Dalrymple, 2001).

References

Allan, Robin, Philip Holmes & Tom Lundskær-Nielsen. 1995. *Danish: A comprehensive grammar*. New York: Routledge.

Askedal, John Ole, Tor Guttu, Per Egil Hegge, Inger-Lise Nyheim, Arthur O. Sandved & Ole Michael Selberg. 2013. *Norsk grammatikk: Riksmål og moderat bokmål*. Oslo: Kunnskapsforlaget.

Barker, Chris. 2000. Definite possessives and discourse novelty. *Theoretical Linguistics* 26. 211–227.

Buss, Samuel R. 1998. An introduction to proof theory. In Samuel R. Buss (ed.), *Handbook of proof theory*, 1–78. Amsterdam: Elsevier.

Crouch, Richard & Josef van Genabith. 1999. Context change, underspecification and the structure of glue language derivations. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 117–189. Cambridge, MA: MIT Press.

Dalrymple, Mary. 1993. *The syntax of anaphoric binding*. Palo Alto, CA: CSLI Publications.

Dalrymple, Mary. 2001. *Lexical Functional Grammar, Syntax and Semantics 34*. New York, NY: Academic Press.

Dipper, Stefanie. 2005. German quantifiers: Determiners or adjectives? In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG05 conference*, 100–115. Stanford, CA: CSLI Publications.

Dotlačil, Jakub. 2012. Binominal *each* as an anaphoric determiner: Compositional analysis. In *Sinn und Bedeutung*, vol. 16.

Dubert, Francisco & Charlotte Galves. 2016. Galician and Portuguese. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to Romance languages*, 411–446. Oxford: Oxford.

Faarlund, Jan Terje. 2004. *The syntax of Old Norse*. Oxford University Press on Demand.

Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14. 39–100.

Heim, Irene. 1982. *On the semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachusetts dissertation.

Heim, Irene. 1997. Predicates or formulas? Evidence from VP-ellipsis. In Aaron Lawon & Eund Cho (eds.), *Proceedings of SALT 7*, 197–221. Ithaca, NY: CLC Publications.

Heim, Irene, Howard Lasnik & Robert May. 1991. Reciprocity and plurality. *Linguistic Inquiry* 22(1).

Hellan, Lars. 1988. *Anaphora in Norwegian and the theory of grammar*. Walter de Gruyter.

Hultman, Tor G. 2003. *Svenska akademiens språklära*. Stockholm: Svenska Akademien.

Kamp, Hans. 1981. A theory of truth and semantic interpretation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, Amsterdam: Mathematisch Centrum, University of Amsterdam.

Link, Godehard. 1983. The logical analysis of plurals and mass terms. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use, and the interpretation of language*, 303–323. Berlin: Walter de Gruyter & Co.

Lødrup, Helge. 2008. Objects binding reflexives in Norwegian. *Norsk lingvistisk tidsskrift* 26(2). 30–54.

Lundquist, Björn. 2014. Binding and co-reference. *Nordic Atlas of Language Structures Journal* 1(1). Http://www.tekstlab.uio.no/nals.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Matthewson, Lisa. 1999. On the interpretation of wide-scope indefinites. *Natural Language Semantics* 7. 79–134.

Milačić, Dejan, Raj Singh & Ida Toivonen. 2015. Distance distributivity and Skolemization. In Eva Csipak & Hedde Zeijlstra (eds.), *Proceedings of Sinn und Bedeutung 19*, 411–425.

Peters, Stanley & Dag Westerståhl. 2013. The semantics of possessives. *Language* 89(4). 713–759.

Platzack, Christer. 1998. *Svenskans inre grammatik: det minimalistiska programmet [The internal grammar of Swedish: The Minimalist Program]*. Lund: Studentlitteratur.

Przepiórkowski, Adam & Agnieszka Patejuk. 2013. The syntax of distance distributivity in Polish: Weak heads in LFG via restriction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of LFG13*, 482–502. Stanford: CSLI.

Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20. 335–397.

Safir, Ken & Tim Stowell. 1988. Binominal each. In *Proceedings of NELS*, vol. 18, 426–450.

Schlenker, Phillippe. 2006. Scopal independence: A note on branching and wide scope readings of indefinites and disjunctions. *Journal of Semantics* 23. 281–314.

Sigurðsson, Halldór Ármann, Einar Freyr Sigurðsson, Alexander Peter Pfaff & Jim Wood. Forthcoming. *Hvor* 'each' reciprocals and distributives in Icelandic. Manuscript.

Spector, Ilona. 2008. *Hebrew floating quantifiers: A non-derivational approach*. Hebrew University of Jerusalem MA thesis.

Steedman, Mark. 2011. *Taking scope: The natural semantics of quantifiers*. Cambridge, MA: MIT Press.

Strunk, Jan. 2004. *Possessive constructions in Modern Low Saxon*. Stanford University MA thesis.

Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik*. Stockholm: Norstedts.

Thráinsson, Höskuldur. 2001. Íslensk setningafrædi. https://notendur.hi.is/hoski/islensksetn.html.

Thráinsson, Höskuldur, Hjalmar Petersen, Jögvan Jacobsen & Zakaris Svabo Hansen. 2004. *Faroese: An overview and reference grammar*. Tórshavn: Føroya Fróðskaparfelag.

Vangsnes, Øystein. 2002a. Distributiv possessiv – en binominal konstruksjon. In Inger Moen, Hanne Gram Simonsen, Arne Torp & Kjell Ivar Vannebo (eds.), *Utvalgte artikler fra det niende møtet om norsk språk i Oslo 2001*, 230–243. Oslo: Novus.

Vangsnes, Øystein. 2002b. Scandinavian binominal constructions: Evidence for a DP-internal analysis, 17th Comparative Germanic Workshop, University of Iceland, ReykjavÍk.

Wechsler, Stephen & Larisa Zlatić. 2000. A theory of agreement and its application to Serbo-Croatian. *Language* 799–832.

Winter, Yoad. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy* 20. 399–467.

Zimmermann, Malte. 2002. *Boys buying two sausages each: On the syntax and semantics of distance-distributivity*. Netherlands Graduate School of Linguistics.

The Argument Structure of Siraiki Causatives

John J. Lowe
University of Oxford

Ali H. Birahimani
Shanghai Jiao Tong University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 191–211

http://csli-publications.stanford.edu/LFG/2019

Keywords: Siraiki, Mapping Theory, causative

Lowe, John J., & Birahimani, Ali H. 2019. The Argument Structure of Siraiki Causatives. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 191–211. Stanford, CA: CSLI Publications.

Abstract

We present data from Siraiki, an understudied modern Indo-Aryan language. Siraiki has two productive morphological causatives, as well as other causative formations. We present the Siraiki data, and develop an analysis within LFG's Mapping Theory. We aim to show that the differences between causativization of intransitive and of transitive verbs can be accounted for without having to assume two homophonous variants of each causative morpheme.

1 Introduction

Siraiki is an understudied modern Indo-Aryan (IA) language spoken in central Pakistan. It has a variety of causative constructions, including both morphological and periphrastic constructions. The details of the Siraiki causative system were previously undescribed, but see now Lowe and Birahimani (2019). In this paper, we focus on the formal analysis of the two most productive causatives, the morphological formations in *-āv* (the 'primary' causative) and *-vāv* (the 'double' causative).

Siraiki is spoken on the western boundary of the IA linguistic area, in the south and west of Pakistan's Punjab province and the southeast of the Khyber Pakhtunkhwa province. It is also the language of a large historically immigrant population in Sindh and of some settlers in the Kacchi region of the Balochistan province. The total number of Siraiki speakers today is likely to be in the range 20–40 million. Siraiki forms a part of a larger dialect continuum with Panjabi toward the east, Sindhi to the south, and Potohari and Hindko in the north. There are significant differences between Siraiki and Panjabi, in phonology and verb system. The Siraiki verb system and phonology are more similar to Sindhi, but in most respects Siraiki and Sindhi are quite different. Beside the reference grammar of Shackle (1976), and brief descriptions in works like Garry and Rubino (2001), there are no detailed recent linguistic accounts or analyses of any linguistic phenomena in Siraiki, before Lowe and Birahimani (2019) and the survey in Bashir and Conners (2019).[1]

The basic grammar of Siraiki is similar to other New IA languages like Panjabi and Hindi-Urdu. It displays two morphological cases, direct vs. oblique, with clitic postpositions for oblique 'cases', and split ergative morphosyntactic alignment. The Siraiki verb system is based on two primary participles, imperfective and perfective. The imperfective participle morpheme (usually) has the form *-(n)d*

[†]We are very grateful to the audiences at LFG19, 8 July 2019, at SE-LFG26, 27 October 2018, and also at Oxford's Syntax Working Group, May 2018 and February 2019, where we made early presentations of this data and analysis. In particular, we thank Rachel Nordlinger and Miriam Butt for helpful discussions. We are also very grateful to Imre Bangha, Ghanshyam Sharma and Peter Hook. All errors are of course our own.

[1]Bashir and Conners's survey is based on a different dialect of Siraiki from that which forms the basis of the present paper and Lowe and Birahimani (2019), and the language data presented there differs in certain respects from our own.

with intransitive verbs, but *-ēnd* with transitive verbs; the imperfective participle with causatives is *-ǣnd*, which represents the contraction of transitive *-ēnd* with the vowel of the causative suffix. In this paper we use only imperfective examples, but our analysis applies equally well to perfective sentences, with the addition of relevant mechanisms to deal with the ergative case marking.

2 Transitive and causative alternations in Siraiki

Dixon (2000) distinguishes four means of expressing causativization: lexical, morphological, analytic, and periphrastic.

Lexical causatives involve a transitivity alternation between two lexically distinct verbs or verb stems. Such alternations are central to the Siraiki verb system, but are distinct from causativization, as shown below. Siraiki has three morphological causatives, formed with the suffixes *-āv*, *-vāv*, and *-āl*; the first two of these are the most productive causative markers in Siraiki, and are the focus of this paper. Siraiki also forms periphrastic causative constructions with *lāvaṇ* and *k^{h}aṛāvaṇ*, for which see Lowe and Birahimani (2019, 286–288).

2.1 Transitivity alternations

A central feature of the Siraiki verb system is the paradigmatic pairing of intransitive and transitive stems. Most verbs display two morphologically distinct but related stems. The largest group of verb pairs involves a vowel alternation whereby the vowel of the transitive stem is underlyingly long and the vowel of the intransitive stem is necessarily short (1a–b). Other morphologically related stems may differ in the final consonant (1c–d), or show both vowel alternation and change in the final consonant (1e–f). Some verb pairs are suppletive (1g–h); a few infrequent verbs, both transitive and intransitive, have no corresponding pair, e.g. transitive *ap^{a}ṛ-* 'catch/seize', *čuṭ-* 'target', *ḍ^{h}ak-* 'cover/imprison', intransitive *baṇ-* 'be built', *b^{h}iṛ-* 'fight'.

(1) a. *ɓal-* 'kindle (intr.)' vs. *ɓāl-* 'kindle (tr.)'
b. *luṛ^h-* 'drift' vs. *lōṛ^h-* 'set adrift'
c. *ʄaṽ-* 'be born' vs. *ʄaṇ-* 'give birth'
d. *d^{h}ōp-* 'be washed' vs. *d^{h}ōv-* 'wash'
e. *vik-* 'be sold' vs. *vēč-* 'sell'
f. *ḍ^{h}uk-* 'be carried' vs. *ḍ^{h}ōh-* 'carry'
g. *ḍ^{h}ah-* 'fall down' vs. *saṭ-* 'throw down'
h. *lab^h-* 'be found' vs. *ɠōl-* 'find'

The following examples (2)–(3) and (4)–(5) serve to illustrate intransitive/transitive verb pairs.[2]

[2]We use standard Leizig glosses, with the addition of DIR for 'direct case', and CS1 and CS2 for the primary and double causative markers, respectively.

(2) *miṭṭī* *ḍ^{h}uk^{a}-d-ī* *ē*
Soil.DIR.F.SG be_carried.INTR-IPFV-F.SG be.PRS.3SG
'The soil is carried.'

(3) *gāmaṇ* *miṭṭī* *ḍ^{h}uh-ēnd-ā* *ē*
Gaman.DIR soil.DIR.F.SG carry.TR-IPFV-M.SG be.PRS.3SG
'Gaman carries the soil.'

(4) *kāṭ^{h}yā̃* *luṛ^{ha}-d-yā̃* *in*
wood.DIR.F.PL drift.INTR-IPFV-F.PL be.PRS.3PL
'The wood drifts.'

(5) *gāmaṇ* *kāṭ^{h}yā̃* *luṛ^{h}-ēnd-ā* *ē*
Gaman.DIR wood.DIR.F.PL set_adrift.TR-IPFV-M.SG be.PRS.3SG
'Gaman sets the wood adrift.'

Transitivity alternations like these are usually treated as a type of causativization, e.g. Dixon (2000), Shibatani and Pardeshi (2002) on Marathi, and most literature on IA languages. But Lowe and Birahimani (2019, 274–276) show that transitivity is grammatically distinct from causativity in Siraiki; consider the following examples.[3]

(6) *gāmaṇ* *niñāṇē̃* *kāṭ^{h}yā̃* *luṛ^{h}-ēnd-ā*
Gaman.DIR unknowingly wood.DIR.F.PL set_adrift.TR-IPFV-M.SG
ē
be.PRS.3SG
'Gaman unknowingly/accidentally sets the wood adrift.'

(7) *sā̃val* **niñāṇē̃* *kāṭ^{h}yā̃* *luṛ^{h}-ǣnd-ā*
Sanwal.DIR unknowingly wood.DIR.F.PL drift.INTR-CS1.IPFV-M.SG
ē
be.PRS.3SG
'Sanwal *unknowingly/accidentally sets the wood adrift.'

In contrast to the subject of the transitive verb (6), the subject of the causative (7) must act intentionally and consciously; adverbs such as *niñāṇē̃* are impossible with the latter. Thus causative formations in Siraiki place particular entailments on their added argument, the *Causer*, which are not found with subjects of ordinary agentive transitive verbs. For further evidence distinguishing transitives from causatives, see Lowe and Birahimani (2019, 274–276).

[3]Transitive stems with underlyingly long vowels show vowel reduction in various forms, including the imperfective, due to stress shift induced by suffixation. Thus *luṛ^{h}-ēnd-ā* is the imperfective of the transitive stem alternant *lōṛ^{h}-* 'set adrift', not of the intransitive *luṛ^{h}-* 'drift' (cf. 1b). On the other hand, the causative *luṛ^{h}-ǣnd-ā* is causative to the intransitive stem (no intermediary causee-agent can be expressed, so this cannot be a causative to the transitive stem with unexpressed causee).

2.2 Causative data

As mentioned above, Siraiki has three morphological causatives, in *-āv*, *-vāv*, and *-āl*. Some verbs show all three, e.g. *sik^h-* 'learn': *sik^{h}āv-* 'make x learn', *sik^{ha}vāv-* 'cause x to make y learn', *sik^{h}āl-* 'teach'. *-āl* is restricted in distribution and will not be considered in this paper, though its analysis would parallel that of *-āv* and *-vāv*.

The morphemes *-āv* and *-vāv* correspond directly with Hindi-Urdu *-ā* and *-vā*, and have cognates in many other IA languages. The 'primary' causative suffix *-āv* derives historically from the (first part of the) Middle Indic causative marker *-āpaya*/*-āpe*/*-āve*.[4] The 'double' causative marker *-vāv* derives from an early double marking *-āpāpaya*/*-āpāpe*/*-āvāve* (found already in the Aśokan inscriptions; Edgerton, 1946, 100), which originally had a double causative value.

2.2.1 The primary causative in *-āv*

The primary causative in *-āv* is older than *-vāv* and up to a certain point in the history of Modern IA was the basic causative suffix with all verbs. In modern Siraiki, it is largely, but not exclusively, restricted to intransitive stems, transitive stems now mostly taking the more productive *-vāv*. The following examples show causativisation of intransitive only *b^{h}iṛ-* 'fight' with the primary causative *-āv* in *b^{h}iṛāv-* 'make x fight':

(8) *murs* *b^{h}iṛ^a-d-ē* *in*
man.DIR fight.INTR-IPFV-M.PL be.PRS.3PL
'Men fight.'

(9) *gāmaṇ* *mursẽ̄=kū̃* *b^{h}iṛ-ǣnd-ā* *ē*
G.DIR man.OBL.M.PL=DAT fight.INTR-CS1.IPFV-M.SG be.PRS.3SG
'Gaman makes men fight.'

The subject of the intransitive verb becomes the object of the causative verb (here differentially marked with *=kū̃* ≈ Hindi-Urdu *-ko*).

Some common transitive stems still form their basic causative with *-āv*, e.g. *ɗik^{h}āv-* 'make x see' from *ɗēk^h-* 'see', *karāv-* 'make x do y' from *kar-* 'do y'. Any analysis of *-āv* must therefore capture the fact that it can attach to both intransitive and transitive stems.

(10) *gāmaṇ=dā* *putr* *kam* *kar-ēnd-ā*
Gaman.OBL=GEN.M.SG.DIR son.DIR work do.TR-IPFV-M.SG
ē
be.PRS.3SG
'Gaman's son does the work.'

[4] The *-aya* element, which was the original causative suffix in Old IA, survives in Siraiki in the *-e-* of the transitive imperfective marker *-ēnd* (and by extension in the causative imperfective marker *-ǣnd*).

(11) *gāmaṇ* *āp^{a}ṇē* *putr=kū̃* *kam* *kar-ǣnd-ā*
Gaman.DIR own.OBL.M.SG son.OBL=DAT work do.TR-CS1.IPFV-M.SG
ē
be.PRS.3SG
'Gaman makes his son do the work.'

The subject of the transitive verb becomes an oblique argument in the causative, marked with the dative case clitic *=kū̃*.[5] The object of the transitive verb remains the object in the causative.

With verbs which have transitive/intransitive stem pairs, the *-āv* causative to the intransitive stem is semantically very close to the transitive stem, but as shown above (6–7) they are distinct.

2.2.2 The 'double' causative *-vāv*

The double causative *-vāv* is the most productive causative suffix in Siraiki. It usually attaches to transitive stems; despite its name (and historical origin), it forms simple (not double) causatives to transitive stems. As when *-āv* attaches to transitive stems, the subject of the transitive becomes an oblique in the causative, while the object of the transitive remains an object. In contrast with *-āv*, the case marking of the oblique causee-agent with *-vāv* causatives is ablative *=dē kanū̃*.[6]

(12) *kap^{a}ṛē* *d^{h}op^{a}-d-ē* *in*
cloth.DIR.M.PL be_washed.INTR-IPFV-M.PL be.PRS.3PL
'The clothes are washed.'

(13) *gāmaṇ=dī* *zāl* *kap^{a}ṛē* *d^{h}u-ēnd-ī*
gaman.OBL=GEN.DIR.F.SG wife.DIR cloth.DIR.M.PL wash.TR-IPFV-F.SG
ē
be.PRS.3SG
'Gaman's wife washes the clothes.'

(14) *gāmaṇ* *zāl=dē* *kanū̃* *kap^{a}ṛē*
gaman.DIR wife.OBL=GEN.OBL.M.SG from cloth.DIR.M.PL
d^{h}u-vǣnd-ā *ē*
wash.TR-CS2.IPFV-M.SG be.PRS.3SG
'Gaman makes his wife wash the clothes.'

The stem *d^{h}u-* in (14) is unambiguously the transitive alternant *d^{h}ōv-* 'wash' (with vowel reduction and loss of *v* after *u*), not the intransitive alternant *d^{h}ōp-*, so this cannot be a double causative to the intransitive stem. Like Hindi-Urdu *-vā*,

[5] Siraiki *=kū̃* therefore has two uses, one for differentially marked direct objects, the other for oblique arguments. The difference is clear under passivization, only the former being promotable to subject.

[6] This difference in case marking can be explained diachronically.

-vāv as a simple causative expresses indirect or non-contactive causation (Saksena, 1982), which is likely to derive historically from its double causative origins.[7]

However, with transitive verbs like *karaṇ* 'to do' which exceptionally admit causatives in *-āv* (16), the causative in *-vāv* does function as a double causative (17):

(15) *gāmaṇ=dā* *putr* *kam* *kar-ēnd-ā*
Gaman.OBL=GEN.M.SG.DIR son.DIR work do.TR-IPFV-M.SG
ē
be.PRS.3SG
'Gaman's son does the work.'

(16) *gāmaṇ* *āpaṇē* *putr=kū̃* *kam* *kar-ǣnd-ā*
Gaman.DIR own.OBL.M.SG son.OBL=DAT work do.TR-CS1.IPFV-M.SG
ē
be.PRS.3SG
'Gaman makes his son do the work.'

(17) *sā̃val* *gāmaṇ=dē* *kanū̃*
Sanwal.DIR Gaman.OBL=GEN.M.SG.OBL from
ū̃=dē *putr=kū̃* *kam* *kara-vǣnd-ā*
DEM.OBL=GEN.M.SG.OBL son.OBL=DAT work do.TR-CS2.IPFV-M.SG
ē
be.PRS.3SG
'Sanwal makes Gaman make his (Gaman's) son do the work.'

The double causative function is also found where *-vāv* attaches to intransitive only stems which also form *-āv* causatives:

(18) *sā̃val* *gāmaṇ=dē* *kanū̃* *mursē̃=kū̃*
sanwal.DIR gaman.OBL=GEN.M.SG.OBL from man.OBL.M.PL=DAT
bʰiṛa-vǣnd-ā *ē*
fight.INTR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal makes Gaman make the men fight.'

(19) *sā̃val* *gāmaṇ=dē* *kanū̃* *saṛkā̃*
sanwal.DIR gaman.OBL=GEN.M.SG.OBL from road.DIR.F.PL
baṇa-vǣnd-ā *ē*
be_built.INTR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal makes Gaman build the roads.'

When the double causative attaches to stems of the *luṛʰ-/lōṛʰ-* type, where vowel reduction neutralizes the distinction between intransitive and transitive stems,

[7] On indirect/non-contactive causation see further the discussion of (25 below.

the analysis is ambiguous; for example, *luṛ*^{ha}*vāvaṇ* 'cause x to set y adrift' could theoretically be double causative from the intransitive verb *luṛ*^{h}*aṇ* 'to drift', or simple causative from transitive verb *lōṛ*^{h}*aṇ* 'to set adrift', or a causative from the primary causative *luṛ*^{h}*āvaṇ* 'to make drift'. This ambiguity may have played a role in the reanalysis of *-vāv* from a double causative marker to a simple causative marker to transitive stems.

(20) *kāṭ*^{h}*yā̃* *luṛ*^{ha}*-d-yā̃* *in*
wood.DIR.F.PL drift.INTR-IPFV-F.PL be.PRS.3PL
'The wood drifts.'

(21) *gāmaṇ* *kāṭ*^{h}*yā̃* *luṛ*^{h}*-ēnd-ā* *ē*
Gaman.DIR wood.DIR.F.PL set_adrift.TR-IPFV-M.SG be.PRS.3SG
'Gaman sets the wood adrift.'

(22) *sā̃val* *kāṭ*^{h}*yā̃* *luṛ*^{h}*-ǣnd-ā* *ē*
Sanwal.DIR wood.DIR.F.PL drift.INTR-CS1.IPFV-M.SG be.PRS.3SG
'Sanwal makes the wood drift.'

(23) *sā̃val* *gāmaṇ=dē* *kanū̃* *kāṭ*^{h}*yā̃*
Sanwal.DIR Gaman.OBL=GEN.M.SG.OBL from wood.DIR.F.PL
luṛ^{ha}*-vǣnd-ā* *ē*
set_adrift.TR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal makes Gaman set the wood adrift.'

Diachronically, at least, *-vāv* represents doubled *-āv*, so conceivably the double causative uses of *-vāv*, such as *kar*^{a}*vāvaṇ* in (17), could be analysed as e.g. *kar-* +*-āv*+*-āv*. However, the more common simple causative use of *-vāv* requires us to treat *-vāv* synchronically as an independent suffix (not merely a contraction of *-āv*+*-āv*).

This simple causative *-vāv* attaches not only to transitive stems, as seen in (14) above, but also to some intransitive stems. For example, besides attaching to the transitive stem *d*^{h}*ōv-* 'wash', it can also attach to the intransitive alternant *d*^{h}*ōp-* 'be washed':

(24) *kap*^{a}*ṛē* *d*^{h}*op*^{a}*-d-ē* *in*
cloth.DIR.M.PL be_washed.INTR-IPFV-M.PL be.PRS.3PL
'The clothes are washed.'

(25) *gāmaṇ* *kap*^{a}*ṛē* *d*^{h}*up*^{a}*-vǣnd-ā*
gaman.DIR cloth.DIR.M.PL be_washed.INTR-CS2.IPFV-M.SG
ē
be.PRS.3SG
'Gaman has the clothes washed.'

Crucially, although the causation in (25) is interpreted as indirect (Gaman does not do the washing himself), it is ungrammatical to add a causee-agent as (26) shows:

(26) **gāmaṇ zāl=dē kanū̃ kap^{a}ṛē*
gaman.DIR wife.OBL=GEN.OBL.M.SG from cloth.DIR.M.PL
d^{h}up^{a}-vǣnd-ā ē
be_washed.INTR-CS2.IPFV-M.SG be.PRS.3SG
'Gaman has his wife wash the clothes.'

Thus (25) cannot represent a double causative with unexpressed causee-agent, but must be a simple (indirect) causative to the intransitive stem.[8]

The pattern seen with *d^{h}ōpaṇ/d^{h}ōvaṇ* is seen also with other verbs, including *ḍ^{h}ukkaṇ/ḍ^{h}ōhaṇ* 'to be carried'/'to carry':

(27) *miṭṭī ḍ^{h}uk^{a}-d-ī ē*
soil.DIR.F.SG be_carried.INTR-IPFV-F.SG be.PRS.3SG
'The soil is carried.'

(28) *gāmaṇ miṭṭī ḍ^{h}uh-ēnd-ā ē*
gaman.DIR soil.DIR.F.SG carry.TR-IPFV-M.SG be.PRS.3SG
'Gaman carries the soil.'

The 'double' causative *-vāv* attaches to *ḍ^{h}ōhaṇ* creating a simple causative (29) to the transitive.

(29) *sā̃val gāmaṇ=dē kanū̃ miṭṭī*
sanwal.DIR gaman.OBL=GEN.M.SG.OBL from soil.DIR.F.SG
ḍ^{h}uh^{a}-vǣnd-ā ē
carry.TR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal makes Gaman carry the soil.'

The intransitive stem *ḍ^{h}uk-* can also form a causative with *-vāv* (30), with the semantic restrictions noted for (25), as seen in (31).

(30) *sā̃val miṭṭī ḍ^{h}uk^{a}-vǣnd-ā ē*
sanwal.DIR soil.DIR.F.SG be_carried.INTR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal has the soil carried.'

(31) **sā̃val gāmaṇ=dē kanū̃ miṭṭī*
sanwal.DIR gaman.OBL=GEN.M.SG.OBL from soil.DIR.F.SG
ḍ^{h}uk^{a}-vǣnd-ā ē
be_carried.INTR-CS2.IPFV-M.SG be.PRS.3SG
'Sanwal has the soil carried by Gaman.'

[8] Note that a primary causative to this stem, expected **d^{h}upāvaṇ*, does not exist.

2.3 Summary

We have shown that Siraiki has two productive morphological causatives, *-āv* and *-vāv*. Although there is a tendency for the former to be found with intransitive stems and the latter with transitive stems, it is crucial for the analysis in the next section that both can and do attach to both intransitive and transitive stems, forming simple causatives in both cases. There are three main differences between *-āv* and *-vāv*: the oblique causee-agent receives different case marking with the two suffixes; the latter can have a double causative sense with some stems (where it is in contrastive distribution with *-āv*); and the latter is also an indirect or non-contactive causative. We address the first two of these issues below; the third we treat as a semantic entailment, the analysis of which goes beyond the argument structure model set up below.

3 Predicate composition in LFG

The argument structure of causatives and similar constructions has a long tradition of analysis within LFG, based on the concept of predicate composition. Important early work was undertaken by Alsina and Joshi (1991) and Alsina (1992, 1996); this was built on and developed most extensively by Butt (e.g. 1995, 1997, 1998, 2014). Much of this work focuses on complex predicates, which require predicate composition in the syntax, but the principles developed are equally well applied to predicate composition in the morphology, as with morphological causatives.

In this paper, we follow Dalrymple et al. (2019) in integrating Butt's approach to complex predication with the model of argument structure developed by Kibort (2001, 2004, 2006, 2007). Kibort's argument structure model has been subject to precise formalization and integration with glue semantics by Findlay (2016), and this has been extended to a glue treatment of complex predication by Lowe (2015, 2019). In this section we present 'traditional' argument structures modelled as complex semantic forms; in section 4 below we reformulate our analysis within the glue-based model of Lowe (2015, 2019).

For reference, Kibort (2007) proposes a universal "valency template" for all non-derived predicates:

(32) $$\left\langle \begin{array}{cccccc} \text{arg}_1 & \text{arg}_2 & \text{arg}_3 & \text{arg}_4 & \ldots & \text{arg}_n \\ [-\textsc{o}/-\textsc{r}] & [-\textsc{r}] & [+\textsc{o}] & [-\textsc{o}] & & [-\textsc{o}] \end{array} \right\rangle$$

Verbs select one or more arg slots together with default feature specification. Arg slots link to grammatical functions according to the hierarchy of grammatical functions: arg_1 (if selected) links to the highest available grammatical function, then arg_2 (if selected) links to the highest remaining grammatical function, and so on. We adopt the hierarchy of grammatical functions proposed by Her (2013), building on Bresnan and Moshi (1990):

(33) $\textsc{subj} > \textsc{obj} > \textsc{obl}_\theta > \textsc{obj}_\theta$

The grammatical functions decompose according to two binary features, ±R(estricted) and ±O(bjective) as proposed by Bresnan and Kanerva (1989). The grammatical function hierarchy can then be understood in terms of markedness: positive values for O and R are marked; markedness is inversely correlated with position on the GF hierarchy.[9]

(34)

	−R	+R
−O	SUBJ	OBL_θ
+O	OBJ	OBJ_θ

3.1 Argument fusion vs. argument raising

Treatments of predicate composition in LFG fundamentally rely on two distinct processes of composition, which Butt (2014) labels 'argument fusion' and 'argument raising'. Alsina and Joshi (1991) and Alsina (1992) first proposed the notion of argument fusion, whereby a causative predicate contains an argument position which is coindexed with an argument position of the embedded predicate. Alongside this, Alsina (1996) also admitted causative predicates which do not show coindexation of arguments, i.e. *argument raising*.

The basic facts of causativisation in Siraiki are naturally similar to the related causatives in other IA languages like Hindi-Urdu, and also to complex predicates in these languages. An analysis of the Hindi-Urdu causative is given by Butt (1997), but most of Butt's work is devoted to Hindi-Urdu complex predicates. For somewhat different reasons from Alsina, Butt (2014) argues that complex predicates in Hindi-Urdu involve both argument fusion and argument raising. Although this is not explicitly extended to causatives by Butt, we assume that in principle the same arguments should hold also for IA causatives. We therefore take the proposals of Butt (2014) to represent the most advanced and up-to-date treatment of predicate composition, applicable also to morphological causatives as found in Siraiki. We briefly discuss the earlier proposals of Butt (1997), which she explicitly applies to the Hindi-Urdu causative, below.

As noted, Butt (2014) argues for two types of causative complex predicate, 'argument fusion' and 'argument raising'. These may be illustrated using her examples of the Urdu permissive, which has two senses: 'allow-to-do' (35) vs. 'allow-to-happen' (36).

(35) *Anjum-ne* *Saddaf-ko* *ciṭṭhii* *likh-ne* *d-ii*
Anjum-ERG Saddaf-DAT note.NOM.F.SG write-INF.OBL let-PFV.F.SG
'Anjum let Saddaf write a note.'

[9] Her (2013) argues that [+R] is more marked than [+O], hence OBJ is higher on the GF hierarchy than OBL_θ.

(36)

kacce	*lamhe-ko*	*ʃaak=par*	*pak-ne*
unripe.M.OBL	moment.M.SG.OBL-ACC	branch.M.SG=on	ripen-INF.OBL

d-o
give-IMP

‘Let the tender moment ripen on the bough.’

Butt (2014) argues that the ‘allow-to-do’ permissive involves ‘argument fusion’: the highest argument of the lexical predicate is coindexed with the lowest non-variable argument of the light verb. For example, the transitive verb *likh-* ‘write’ has two arguments:

(37)

Saddaf-ne	*ciṭṭhii*	*likh-ii*
Saddaf-ERG	note.NOM.F.SG	write.PFV.F.SG

‘Saddaf wrote a note.’

Stated in terms of Kibort’s (2007) valency template, the argument structure for *likh-*, and its basic mapping to grammatical functions, can be represented as follows:

(38)

		AGENT	THEME	
		|	|	
likh ‘write’	⟨	arg_1	arg_2	⟩
		[−O]	[−R]	
		SUBJ	OBJ	

The light verb *de-* has a three place argument structure, including a variable %PRED, which is filled by the embedded predicate:

(39) ‘allow-to-do’ *de*: ‘let⟨arg_1, arg_3, %PRED⟩’

In analysing *likhne de-* in (35), the argument frame for *likh-* is embedded in the argument frame for *de-*, and argument fusion coindexes the arg_3 of *de* with the arg_1 of *likh-*:

(40)

de ‘let’	⟨	arg_1	arg_3	‘write’	⟨	arg_1	arg_2	⟩⟩
		[−O]	[+O]			([−O])	[−R]	
		SUBJ	OBJ_θ				OBJ	

The coindexed argument adopts its feature specification from the arg_3 of the light verb, and hence links to OBJ_θ.

Butt (2014) argues that the ‘allow-to-happen’ permissive in (36) involves argument raising rather than argument fusion. The verb *pak-* ‘ripen’ is intransitive and has only one argument position. In this case, the light verb *de-* ‘let’ is assumed to have a different subcategorisation frame, with one argument position plus the variable slot %PRED:

(41) 'allow-to-happen' *de*: 'let⟨arg_1, %PRED⟩'

In the formation of the complex predicate, the subcategorisation frame of *pak-* is simply inserted into that of the light verb.

(42)
$$\begin{array}{lccccccc} \textit{de}\ \text{'let'} & \langle & \text{arg}_1 & \text{'ripen'} & \langle & \text{arg}_1 & \rangle\rangle \\ & & [-\text{O}] & & & [-\text{R}] & \\ \hline & & \text{SUBJ} & & & \text{OBJ} & \end{array}$$

The arg_1 of the embedded predicate retains its [−R] specification, and therefore links to OBJ.

We must consider why two different types of predicate composition are required to deal with the Hindi-Urdu permissive (the same considerations would apply to an analysis of the Siraiki causatives along these lines). As we see it, there is one practical (syntactic) argument, and one conceptual (semantic) argument. The latter will be discussed in the next section. On the syntactic side, there is a fundamental difference between predicate composition applied to a transitive verb and predicate composition applied to an intransitive verb: in the former case, the subject (or arg_1) of the embedded predicate becomes a restricted argument (OBJ_θ in the case of the Hindi-Urdu permissive, OBL_θ for the Siraiki causative), but in the latter case the subject/arg_1 of the embedded predicate becomes the OBJ of the resulting composed predicate. By distinguishing argument fusion from argument raising, and using one for composition with transitive predicates and the other for composition with intransitive predicates, this difference can be resolved.[10]

It is worth noting that there are alternatives to this argument fusion/raising distinction as a means of accounting for the differential treatment of the arg_1 of transitive/intransitive verbs. In the treatment of Hindi-Urdu causatives proposed in Butt (1997), only argument fusion is required, but crucially the causative morpheme still has two different argument structures, to account for the different realizations of the embedded arg_1. The same is true of the XLE approach (see Lowe, 2015). Under the rather different proposals of Alsina (1996), the difference between the two could be captured by assuming that with transitive verbs, argument fusion coindexes the *lowest* argument of the embedded predicate, rather than the highest. But Butt (1997) argues strongly against such a possibility, proposing the 'Restriction on Argument Fusion': "Only the highest θ-rule may escape its domain of predication, and thus become eligible for Argument Fusion."

Under any previous analysis, then, two homophonous versions of any causative (or permissive, etc.) are required to deal with causativization of transitive and intransitive verbs. We argue that this need for two homophonous causative predicates is far from ideal: there are obvious differences between causativization of intransitive and transitive verbs, but ideally these ought to fall out directly from the different properties of transitive and intransitive verbs, without requiring two homophonous variants of every causative morpheme. For Siraiki causatives, the

[10]For a problematic exception to this generalization, see §3.4 below.

problem is particularly acute, since as we have shown above, both *-āv* and *-vāv* can be used with both transitive and intransitive verbs, with the same argument structure patterns in both cases; therefore we would have to assume that both show exactly the same ambiguity. In fact, the third Siraiki causative morpheme *-āl* can also causativize both intransitive and transitive verbs, so we would end up having to assume the same ambiguity for three causative morphemes. It seems like this would be missing a generalization.

3.2 Our proposal

We propose that in Siraiki, and also in Hindi-Urdu (and probably many other IA languages), we in fact only need one argument frame for causativization (and similar processes, like the permissive), which derives the required differences between transitive and intransitive attachment entirely through the properties of the embedded predicate. This frame involves argument fusion, but crucially the second argument position is *unspecified for* [±O/R] *features*. Argument raising is thus not required.

(43)
$$
\begin{array}{llcccc}
 & & \text{CAUSER} & \text{CAUSEE} & & \\
 & & | & | & & \\
\textit{-āv}\ \text{‘CAUSE’} & \langle & \text{arg}_1, & \text{arg}_4, & \%\text{PRED} & \rangle \\
 & & [-\text{O}] & [\,] & &
\end{array}
$$

This single argument frame can be applied to both intransitive and transitive predicates, deriving the correct argument structure for the respective causatives. Example (44) shows the monovalent argument structure of the intransitive verb *b^hiṛ-* ‘fight’, from example (8); ex. (45) shows the complex argument structure of the causative *b^hiṛ-āv-* ‘make x fight’ from example (9).

(44)
$$
\begin{array}{lccc}
\textit{b}^{h}\textit{iṛ}\ \text{‘fight’} & \langle & \text{arg}_1 & \rangle \\
 & & [-\text{O}] &
\end{array}
$$

(45)
$$
\begin{array}{llcccccc}
 & & \text{CAUSER} & \text{CAUSEE} & & & \text{AGENT} & \\
 & & | & | & & & | & \\
\textit{-āv}\ \text{‘CAUSE’} & \langle & \text{arg}_1 & \text{arg}_4 & \textit{b}^{h}\textit{iṛ}\ \text{‘fight’} & \langle & \text{arg}_1 & \rangle\rangle \\
 & & [-\text{O}] & [\,] & & & ([-\text{O}]) & \\
\hline
 & & \text{SUBJ} & \text{OBJ} & & & &
\end{array}
$$

Note that it is always the properties of the outer/higher fused argument which are relevant for linking. The linking to SUBJ and OBJ proceeds without problem. The same causative predicate applied to a transitive verb is slightly more complicated; the following argument structures represent transitive *kar-* ‘do’ and causative *kar-āv-* ‘make x do’ from (10) and (11) respectively.

(46)

		AGENT	THEME	
kar 'do'	⟨	arg_1,	arg_2	⟩
		[−O]	[−R]	
		SUBJ	OBJ	

(47)

		CAUSER	CAUSEE			AGENT	THEME	
-*āv* 'CAUSE'	⟨	arg_1	arg_4	*kar* 'do'	⟨	arg_1	arg_2	⟩⟩
		[−O]	[ ]			([−O])	[−R]	
		SUBJ	OBL_θ				OBJ	

Kibort (2007, 265) proposes a unified Mapping Principle, whereby "the ordered arguments [of a predicate, ordered according to the valency template] are mapped onto the highest (i.e. *least* marked) compatible function on the markedness hierarchy." This considers only simple argument structure, and leaves a degree of ambiguity regarding complex structures like that in (47). In simple cases, the linear (left-to-right) ordering of arguments always matches the numerical ordering of the arg indices, based on the valency template given in (32). In (47), however, the arg_4 of the causative predicate linearly precedes the arg_2 of the embedded predicate. We propose that in such cases, at least in Siraiki, ordering of arguments for linking purposes is based on arg index, not linear order.[11] In the case of (47), then, after the arg_1 of the causative predicate links to SUBJ, the arg_2 of the embedded predicate links to the highest remaining compatible grammatical function, OBJ, and finally the arg_4 of the causative predicate links to the highest remaining compatible function, OBL_θ. If linking were to proceed according to linear order, the arg_2 of the causative predicate would link to OBJ, and the arg_4 of the embedded predicate would then be unable to link, as there would be no [−R] functions remaining.

The unspecified arg_4 argument position does not appear in Kibort's (2007) valency template; we must therefore assume an augmented version of this template. Below, we will argue for a second unspecified position for the -*vāv* causative; we therefore assume the following revised version of Kibort's valency template:

(48) New proposed valency template:

⟨	arg_1	arg_2	arg_3	arg_4	arg_5	arg_6	…	arg_n ⟩
	[−O/−R]	[−R]	[+O]	[ ]	[ ]	[−O]		[−O]

The conceptual argument for the argument fusion/raising distinction in Butt's model, mentioned above, involves the status of the causee (or permittee): there is a difference between permitting/causing someone to do something, and permitting/causing something to happen ('allow-to-do' vs. 'allow-to-happen'). Argument fusion in the former case captures the fact that the act of permitting/causing someone to do something distinctly involves an affected permittee/causee, whereas permitting/causing something to happen less distinctly involves a permittee/causee as

[11] But see Lowe et al. (2019) for Sanskrit data where ordering based on linearity is required.

such. That is, for Butt (2014), Saddaf in (35) is directly involved in the act of permission in a way that the tender moment in (36) is not. This is best understood as a claim about the semantic entailments holding on the permittee/causee role: a permittee or causee must be directly affected by the act of permission/causation, and also sufficiently agentive to bring about the embedded event. Our proposal, which involves argument fusion for all causatives, cannot claim to capture such differences in the argument structure. Our claims about the semantic entailments on the causee are less strict, and as such it is unproblematic to assume that the subjects of both transitive and intransitive verbs can be causees. This is certainly what is implied, in any case, by the analyses of causatives by Alsina (1996) and Butt (1997), so we assume that a relatively noncommittal approach to the semantic entailments on causees is not unjustified.[12]

3.3 *-vāv* and the double causative

Exactly the same type of argument frame will work for the simple causative uses of *-vāv* (and *-āl*), again whether attached to transitive or intransitive stems; the different case marking of the oblique argument can be specified separately (see below), and the non-contactive nature of the causation implied by *-vāv* is a semantic entailment which could for example be represented at semantic structure.

However, the double causative use of *-vāv* requires additional analysis. As seen in (17), when *-vāv* has a double causative value and both intermediate arguments are expressed, the 'causee-causer' (the causee of the 'outer' causation event and the causer of the 'inner' causation event) has the ablative case marking associated with the oblique argument of *-vāv*, while the 'causee-agent' (the causee of the 'inner' causation event and the agent of the lexical predicate) has the dative case marking associated with the oblique argument of *-āv*. Thus *-vāv* in double causative value functions as if it were synchronically *-āv*+*-vāv*.

In fact, this is not unexpected: Kulikov (1993, 126) notes a typological pattern of 'double affix reduction' whereby double causative marking can be reduced to marking by a single causative morpheme without changing the double causative sense. We therefore propose that the double causative use of *-vāv* represents the surface realization of what is functionally a double suffixation *-āv-vāv*.[13]

In this case, then, *-āv* and *-vāv* co-occur in the same verb form, underlyingly. In order that there be no clash between the two suffixes, particularly in terms of their respective case marking specifications (and with a view to the glue analysis below), we now require two distinct unspecified arg slots: unspecified arg_4 for *-āv* and unspecified arg_5 for *-vāv*.

[12]We have considered in depth whether there could be any evidence whatsoever in Siraiki regarding the semantic entailments on causees: if there were some way to distinguish the entailments placed on the embedded arg_1 of transitive verbs from those placed on the embedded arg_1 of intransitive verbs, this could provide evidence for the fusion/raising distinction. We have been unable to identify any means of distinguishing these arguments, however, which correlates with, but cannot be taken to positively support, our approach.

[13]We left this question open in Lowe and Birahimani (2019).

(49) *sā̃val* *gāmaṇ=dē* *kanū̃*
Sanwal.DIR Gaman.OBL=GEN.M.SG.OBL from
ū̃=dē *putr=kū̃* *kam* *kar^{a}-vænd-ā*
DEM.OBL=GEN.M.SG.OBL son.OBL=DAT work do.TR-CS2.IPFV-M.SG
ē
be.PRS.3SG
'Sanwal makes Gaman make his (Gaman's) son do the work.'

(50)

	CAUSER	CAUSEE		CAUSER	CAUSEE		AGENT	THEME
	|	|		|	|		|	|
-vāv	⟨ arg_1	arg_5	⟨⟨ *-āv*	⟨ arg_1	arg_4 ⟩	*kar* 'do'	⟨ arg_1	arg_2 ⟩⟩⟩
	[−O]	[ ]		([−O])	[ ]		([−O])	[−R]
	SUBJ	OBL_θ			OBL_θ			OBJ

3.4 A problem in Hindi-Urdu

Above we presented the proposals of Butt (2014) in terms of a distinction between transitive and intransitive verbs: Butt (like previous proposals) requires two distinct argument frames for the same causative (or permissive) predicate in order to account for the different treatment of the embedded first argument of intransitive and transitive verbs respectively. The reality is slightly more complex: Butt (2014, 186–187) gives an example (based on Davison, 2014, 141) of a permissive of an intransitive verb where the permittee becomes OBJ_θ rather than OBJ. In Butt's analysis, OBJ_θ is the grammatical function of the embedded first argument of a transitive verb (equivalent to our OBL_θ for Siraiki causatives), but in this case is found with the first argument of an intransitive verb.

(51) *mā̃-ne* *bacce-ko* *jā-ne* *dī-yā*
mother-ERG child.OBL-DAT go-INF give.PFV.M.SG
'The mother let the child go.'

Here, according to Davison and Butt, the dative marking on 'child' marks the indirect object, not a differentially marked direct object.[14] That is, the permissive

[14] For the sake of argument, we accept this analysis here. However, there are complications with Davison's claim that *bacce-ko* here must be an indirect object. It is possible to have an inanimate direct object with the permissive of *jānā*, as shown in the following example, suggesting that *bacce-ko* may show DOM.

(1) *caokīdār-ne* *gāṛī* *jā-ne* *dī*
guard-ERG car go-INF give.PFV.F.SG
'The guard let the car go.'

Moreover, Davison's argument rests on the fact that dative marking on 'child' is obligatory in the passive of (51), whereas in the passives of transitive verbs generally the nominative/direct case is also possible for original DOM objects which are promoted to subject. However, we believe that this obligatory dative marking may reflect the semantic similarity of the passive permissive to certain impersonal modal constructions, such as the nominative infinitive + *cāhiye* expressing the sense of some intransitive verbs in Hindi-Urdu treats the first (single) argument of the embedded predicate in the same way as the first argument of an embedded transitive predicate.

This type does not occur in Siraiki causatives, where the treatment of intransitive verbs is consistently as presented above. In order to extend our proposals to Hindi-Urdu, however, we would need to be able to deal with this type. We must first note that it is not entirely clear how the correct analysis derives in Butt's model. The following example gives the analysis of Butt (2014, 187, ex. 51), (trivially) reformulated in terms of Kibort's valency template:

(52)
$$\begin{array}{cccccccc} & & \text{AGENT} & \text{GOAL} & & & \text{THEME} & \\ & & | & | & & & | & \\ \text{GIVE/LET} & \langle & \text{arg}_1 & \text{arg}_3 & \text{GO} & \langle & \text{arg}_1 & \rangle\rangle \\ & & [-\text{O}] & [+\text{O}] & & & ([-\text{R}]) & \\ \hline & & \text{SUBJ} & \text{OBJ}_{\text{goal}} & & & & \end{array}$$

Butt (2014) assumes that the [+O] specification on the second argument of the permissive is sufficient to guarantee the restricted object function, but it is equally possible that [+O] could link to OBJ, and given Kibort's (2007) Mapping Principle, OBJ would be the only possibility in this case, as it is the highest, least marked, available argument on the grammatical function hierarchy after arg_1 links to SUBJ. Butt (2014) does not explain how OBJ is avoided in her analysis.

In the terms of our analysis above, we can analyse this type by means of a specification on the lexical verb: this type is lexically restricted to a set of intransitive verbs. Such verbs may contain a specification determining how they combine with certain complex predicates. In our terms, such a specification in this case would state that if there is an arg_4 in the argument frame of (or including) the verb, that arg_4 gets the additional specification [+R]:[15]

(53)
$$\begin{array}{cccccccc} & & \text{PERMITTER} & \text{PERMITTEE} & & & \text{THEME} & \\ & & | & | & & & | & \\ \text{GIVE/LET} & \langle & \text{arg}_1 & \text{arg}_4 & \text{GO} & \langle & \text{arg}_1 & \rangle\rangle \\ & & [-\text{O}] & [\ ] & & & ([-\text{R}]) & \\ \text{Lex.Spec.} & & & [+\text{R}] & & & & \\ & & \text{SUBJ} & \text{OBL}_{\text{goal}} & & & & \end{array}$$

4 The glue approach

Lowe (2015, 2019) discusses various problems with the standard LFG (and XLE) account of complex predicates assumed here, and proposes an alternative glue-based analysis within Asudeh and Giorgolo's (2012) approach to argument alternations. As given, Lowe's account would also require multiple argument frames for causative morphemes.

'should', which obligatorily show dative marking on animate subjects.

[15]This is easily formalized in the glue model discussed below.

However, it is simple to reformulate the approach proposed here within Lowe's approach, if we adopt the argument structure implementation of Findlay (2016). The causative morpheme introduces the following functional descriptions and meaning constructor (very similar to Lowe, 2015, 434); given the relevant principles of mapping (cf. Findlay, 2016, 322), this will combine unproblematically with the f-descriptions and meaning constructors introduced by intransitive and transitive verbs, permitting a single analysis of both.

(54) *-āv* 'CAUS'

$$(\uparrow \{\text{SUBJ}|\text{OBL}_\theta\})_\sigma = (\uparrow_\sigma \text{ ARG}_1)$$
$$(\uparrow \{\text{SUBJ}|\text{OBJ}|\text{OBL}_\theta|\text{OBJ}_\theta\})_\sigma = (\uparrow_\sigma \text{ ARG}_4)$$

$$\lambda P.\lambda y.\lambda x.\lambda e.caus(x, y, P(y, e)) :$$
$$[(\uparrow_\sigma \text{ARG}_1) \multimap (\uparrow_\sigma \text{EV}) \multimap \uparrow_\sigma] \multimap$$
$$(\uparrow_\sigma \text{ARG}_4) \multimap (\uparrow_\sigma \text{ARG}_1) \multimap (\uparrow_\sigma \text{EV}) \multimap \uparrow_\sigma$$

The equivalent will apply for *-vāv*, with arg_5 for arg_4 in all occurrences. Essentially, the only difference here from Lowe (2015, 434) is that the arg_4 is entirely unspecified as to its associated grammatical function, permitting it to link to OBJ or OBL_θ as required.

We have not formalized the case marking requirements of the causative predicates in our analysis above, but this can be done unproblematically within the glue model of Lowe (2015, 2019). We assume that the casemarking of the core arguments, subject and object, is subject to general specifications: direct case for subjects of intransitives, subjects of transitive in imperfective aspect, and (≈)inanimate objects of transitive verbs; ergative case for subjects of transitives in the perfective; dative for differentially marked (≈animate) objects of transitive verbs.

Given these general specifications, the causative morphemes introduce default specifications for the case of their 'causee' arguments:

(55) a. *-āv*: $\{((\uparrow_\sigma \text{ ARG}_4)_{\sigma^{-1}}\text{CASE})|((\uparrow_\sigma \text{ ARG}_4)_{\sigma^{-1}}\text{CASE})\} = \text{DAT}$
b. *-vāv*: $\{((\uparrow_\sigma \text{ ARG}_5)_{\sigma^{-1}}\text{CASE})|((\uparrow_\sigma \text{ ARG}_5)_{\sigma^{-1}}\text{CASE})\} = \text{ABL}$

Informally, these specifications state that either the causee argument has case specified by something else, or else its case is DAT/ABL.

5 Conclusion

Siraiki has a rich causative system, with two productive morphological causatives as well as another morphological causative and periphrastic causative expressions. All three morphological causatives, of which in this paper we have focused on the two most productive, can attach to both intransitive and transitive verbs. The existing LFG standard for dealing with complex predication of this sort would involve assuming two homophonous variants of each causative suffix, in order to deal with the differential treatment of intransitive and transitive verbs, but we have argued that such an approach misses a generalization. We have shown that a slightly different approach is possible, building on existing work but assuming a single causative argument frame for causatives of both transitive and intransitive verbs.

Causatives and related formations are highly complex and varied, both in Siraiki and crosslinguistically, and our account is necessarily restricted in some respects. Future work could profitably explore extending this approach to other phenomena, both in Siraiki and other modern Indo-Aryan languages, and more widely. For an initial step in this direction, see Lowe et al. (2019).

References

Alsina, Alex (1992). 'On the argument structure of causatives'. *Linguistic Inquiry 23*(4), pp. 517–555.

Alsina, Alex (1996). *The Role of Argument Structure in Grammar*. Stanford, CA: CSLI Publications.

Alsina, Alex and Smita Joshi (1991). 'Parameters in Causative Constructions'. In L. Dobrin, L. Nichols, and R. M. Rodriguez (eds.), *Proceedings from the 27th Regional Meeting of the Chicago Linguistic Society (CLS)*, Chicago, IL: Chicago Linguistic Society, pp. 1–16.

Asudeh, Ash and Gianluca Giorgolo (2012). 'Flexible Composition for Optional and Derived Arguments'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, Stanford, CA: CSLI Publications, pp. 64–84.

Bashir, Elena and Thomas J. Conners (2019). *Grammar of Hindko, Panjabi, and Saraiki*. Berlin: de Gruyter.

Bresnan, Joan and Jonni M. Kanerva (1989). 'Locative Inversion in Chicheŵa: A Case Study of Factorization in Grammar'. *Linguistic Inquiry 20*, pp. 1–50.

Bresnan, Joan and Lioba Moshi (1990). 'Object Asymmetries in Comparative Bantu Syntax'. *Linguistic Inquiry 21*(2), pp. 147–185.

Butt, Miriam (1995). *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications.

Butt, Miriam (1997). 'Complex Predicates in Urdu'. In Alex Alsina, Joan Bresnan, and Peter Sells (eds.), *Complex Predicates*, Stanford, CA: CSLI Publications.

Butt, Miriam (1998). 'Constraining Argument Merger Through Aspect'. In Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, San Diego, CA: Academic Press, pp. 73–113.

Butt, Miriam (2014). 'Control vs. Complex Predication: Identifying Non-Finite Complements'. *Natural Language & Linguistic Theory 32*(1), pp. 165–190.

Dalrymple, Mary, John J. Lowe, and Louise Mycock (2019). *The Oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Davison, Alice (2014). 'Non-finite complements and modality in *de-na* 'allow' in Hindi-Urdu'. *Natural Language & Linguistic Theory 32*, pp. 137–164.

Dixon, R. M. W. (2000). ‘A typology of causatives: form, syntax and meaning’. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds.), *Changing valency: Case studies in transitivity*, Cambridge: Cambridge University Press, pp. 30–83.

Edgerton, Franklin (1946). ‘Indic causatives in *-āpayati* (*-āpeti*, *-āvei*)’. *Language 22*(2), pp. 94–101.

Findlay, Jamie Y. (2016). ‘Mapping Theory without Argument Structure’. *Journal of Language Modelling 4*(2), pp. 293–338.

Garry, Jane and Carl R. Galvez Rubino (2001). *Facts about the world’s languages: an encyclopedia of the world’s major languages, past and present*. New York: The H.W. Wilson Company.

Her, One-Soon (2013). ‘Lexical Mapping Theory Revisited’. In Tracy Holloway King and Valeria de Paiva (eds.), *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, Stanford, CA: CSLI Publications, pp. 47–59.

Kibort, Anna (2001). ‘The Polish passive and impersonal in Lexical Mapping Theory’. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, Stanford, CA: CSLI Publications, pp. 163–183.

Kibort, Anna (2004). *Passive and Passive-Like Constructions in English and Polish*. Ph.D. thesis, University of Cambridge.

Kibort, Anna (2006). ‘On Three Different Types of Subjectlessness and How to Model Them in LFG’. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna (2007). ‘Extending the Applicability of Lexical Mapping Theory’. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, Stanford, CA: CSLI Publications, pp. 250–270.

Kulikov, Leonid (1993). ‘The “second causative”: A typological sketch’. In Bernard Comrie and Maria Polinsky (eds.), *Causatives and transitivity*, Amsterdam: Benjamins, pp. 121–154.

Lowe, John J. (2015). ‘Complex predicates: an LFG+Glue analysis’. *Journal of Language Modelling 3*(2), pp. 413–462.

Lowe, John J. (2019). ‘Argument alternations in complex predicates: an LFG+glue perspective’. *Linguistic issues in Language Technology 17*(2), pp. 1–16.

Lowe, John J. and Ali H. Birahimani (2019). ‘Causative alternations in Siraiki’. *Transactions of the Philological Society 117*(2), pp. 266–293.

Lowe, John J., Adriana Molina-Muñoz, and Antonia Ruppel (2019). ‘Causatives and their passives in Sanskrit’. In Miriam Butt, Tracy Holloway King, and Ida Toivonen (eds.), *Proceedings of the LFG19 Conference*, Stanford, CA: CSLI.

Saksena, Anuradha (1982). ‘Contact in causation’. *Language 58*, pp. 820–831.

Shackle, Christopher (1976). *The Siraiki language of central Pakistan: A reference grammar*. London: School of Oriental and African Studies, University of London.

Shibatani, Masayoshi and Prashant Pardeshi (2002). ‘The causative continuum’. In Masayoshi Shibatani (ed.), *The grammar of causation and interpersonal manipulation*, Amsterdam: Benjamins, pp. 85–126.

Causatives and their Passives in Sanskrit

John J. Lowe
University of Oxford

Adriana Molina-Muñoz
University of Oxford

Antonia Ruppel
University of Oxford

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 212–232

`http://csli-publications.stanford.edu/LFG/2019`

Keywords: causative, passive, Sanskrit

Lowe, John J., Molina-Muñoz, Adriana, & Ruppel, Antonia. 2019. Causatives and their Passives in Sanskrit. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 212–232. Stanford, CA: CSLI Publications.

Abstract

We discuss data on causative constructions and their passives in Sanskrit. Sanskrit is unusual in licensing two different causative constructions for most verbs, together with corresponding passives. We explore the formal analysis of these patterns in argument structure terms, formalized within the argument structure proposals of Kibort (2007) and the approach to causative argument structure proposed by Lowe and Birahimani (2019).

1 Introduction

Sanskrit is an old Indo-Aryan language which was originally spoken in the Northwest of the Indian subcontinent in the first and second millennia BC; it ceased to be a living language during the first millennium BC, but its importance grew, as it became a lingua franca, especially for academic, literary, and religious discourse, throughout India in the first and second millennia AD. Classical Sanskrit has a productive morphological causative, which shows two possible argument structure patterns with most transitive verbs. Both possible patterns are found in other languages, and constitute the two main ways in which languages form causatives to transitives, but relatively few languages have been described as freely permitting both possibilities alongside one another. In addition, Sanskrit has a highly productive morphological passive, which can be applied to any causative verb. The use of passive causatives has been claimed to be more frequent in Sanskrit than in other languages in which this combination occurs, due to the high productivity and use of the passive voice (Bubeník 1987). Sanskrit passives of causatives are further interesting in being able to passivize on either the original subject or object of the base predicate, corresponding to the two patterns of active causative to transitive verbs.

Licensing both causative patterns, and their corresponding passives, for a single causative morpheme, proves problematic for existing LFG analyses of predicate composition within Kibort's (2007) approach to (Lexical) Mapping Theory. In §2 we present the Sanskrit data on causatives and their passives. In §§3–5 we present our formal analysis, discussing three possible ways of analysing the data within a mapping theory approach to complex predicates. In §6 we draw conclusions.

2 Data

The basic possibilities for causativization, passivization, and their combination, in Sanskrit have been known for a long time, see e.g. the overviews in Speyer (1886: 32–38) and Renou (1961: 472–473). Kiparsky and Staal (1969) provide an early

[†]We are grateful to all those who visited our poster at LFG19, and in particular to Miriam Butt, Fengrong Yang, Rigardt Pretorius, Ansu Berg and K Sarveswaran for helpful comments and suggestions. All errors are our own. This work is part of the project 'Uncovering Sanskrit Syntax', funded as a Research Project Grant (RPG-2018-157) by the Leverhulme Trust.

generative treatment of some of the patterns, with a focus on the prescriptions of the authoritative native grammarian Pāṇini. Cardona (1978), Hock (1981), Jamison (1983) and Kulikov (2013) provide detailed treatments of causativization and its origins in the earliest attested stage of Sanskrit, Vedic; Kulikov (2012) includes discussion of passive causatives attested in Vedic. Deshpande (1991) discusses some related patterns in passivization with ditransitive verbs. The existence of two parallel causative argument structures for transitive verbs is mentioned in passing by Aissen (1979: 16–17, 78), but without analysis. The only detailed treatment of passives of causatives in Classical Sanskrit is by Bubeník (1987), who however focuses on data from a very small corpus. Our observations on attested patterns below are based on a new large-scale corpus investigation.[1]

2.1 Causativization

In Sanskrit, causativization is a morphological operation involving a suffix *-áya-* and various largely predictable ablaut alternations in the verbal root. Thus, for instance, to the verbal root $\sqrt{}$*hṛ* 'to take' we get the simple present *har-a-ti* 'takes' and the causative present *hār-aya-ti* 'causes to take'; to the root $\sqrt{}$*viś* 'to enter' we get the simple present *viś-a-ti* 'enters' and the causative present *veś-aya-ti* 'causes to enter'.

With most verbs, this process is regular and productive; with some verbs, including some common verbs, the causative can have an unpredictable idiomatic or lexicalized meaning: *ākārayati*, causative of *ā-*$\sqrt{}$*kṛ* (*ā-* 'hither'; $\sqrt{}$*kṛ* 'do, make') means 'to invite'; *darśayati*, causative of $\sqrt{}$*dṛś* 'to see', means 'show' (something to someone) rather than 'cause (someone) to see (something)'. With several transitive verbs, the causative is semantically indistinguishable from its corresponding basic verb: thus, the simple present and the causative of $\sqrt{}$*vṛ* (*vṛṇoti* and *vārayati*, respectively) both mean 'to cover'.[2]

In this paper we focus only on the semantically and morphosyntactically regular causatives, and restrict ourselves largely to the more interesting case of causatives of transitive verbs, whose argument structures involve three arguments: the two original arguments of the uncausativized (base) predicate, and an added causer.

When an intransitive verb is causativized, the original subject of the base predicate becomes the object of the resulting predicate (marked with the accusative

[1]The corpus consists of ca. 5 million words, spans over two millennia and includes all major Sanskrit textual genres. For reasons of space we cannot provide further details here.

[2]In many cases this is an accidental reflex of earlier historical developments, but it also becomes an ongoing feature of the *-áya-* causative that spreads in Middle Indic, with the result that the direct reflex of the *-áya-* causative suffix in modern Indo-Aryan (where it survives) is as a simple transitivity marker; see Bloch (1965: 239–242) and Masica (1991: 315–321) for the historical development of causative suffixes in Indo-Aryan. Butt (1998: §4.1) incorrectly states that *-áya-* underlies Hindi-Urdu *-ā* (and related causative morphemes in other modern IA languages); in fact, *-ā* comes from early Indo-Aryan *-āpaya-*, while Hindi-Urdu *-vā* (and related morphemes) come from Middle Indo-Aryan doubled causatives in *-āpāpaya-*.

case); this is illustrated in (1).[3]

(1) a. *devadattaḥ śete*
D.NOM lie.PRS.3SG
'Devadatta lies down.'
b. *yajñadatto devadattaṃ śāyayati*
Y.NOM D.ACC lie.CAUS.PRS.3SG
'Yajñadatta makes Devadatta lie down.'

When a transitive verb is causativized, there are, for most verbs, two possibilities. These possibilities correspond to the two main strategies for causativization of transitive verbs attested cross-linguistically. For example, Baker (1988: 161–167) proposes two "causative rules" found in different languages; these differ crucially in the treatment of the original subject of a transitive base predicate. Under "causative rule 1", the subject of an original transitive verb surfaces as an oblique or indirect object in the causative, whereas under "causative rule 2" the subject of an original transitive verb surfaces as an object in the causative.

As stated, most verbs in Sanskrit can show both these possibilities in the causative. The sentences in (2) illustrate the second type: the subject of the base predicate appears as the object in the causative (marked in accusative case), with the object of the base predicate also appearing in the accusative. We call this the accusative-accusative (ACC-ACC) type.

(2) a. *devadatto kaṭaṃ karoti*
D.NOM mat.ACC make.PRS.3SG
'Devadatta makes a mat'
b. *yajñadatto devadattaṃ kaṭaṃ kārayati*
Y.NOM D.ACC mat.ACC make.CAUS.PRS.3SG
'Yajñadatta makes Devadatta make a mat.'

Alternatively, the original subject of the base predicate becomes an oblique argument (usually marked with instrumental case), with the original object of the base predicate remaining the object of the causative; this is illustrated in (3).

(3) *yajñadatto devadattena kaṭaṃ kārayati*
Y.NOM D.INSTR mat.ACC make.CAUS.PRS.3SG
'Yajñadatta makes Devadatta make a mat.'

We call this the 'oblique-accusative' (OBL-ACC) type.[4] As noted above, the ACC-ACC and OBL-ACC causatives correspond to the two major strategies of causativization attested cross-linguistically. However, most languages show only one

[3]The examples are our own, based on real examples found in the Sanskrit corpus and on those discussed by the ancient indigenous grammatical tradition.

[4]The instrumental case marking is primarily semantic, marking agency. With experiencer verbs like $\sqrt{}$*jñā* 'to know' and $\sqrt{}$*śru* 'to hear', the expected semantic case, dative/genitive, almost always occurs in place of the instrumental, but some examples occur where the instrumental is used in place of a semantically more appropriate case, showing a degree of syntactic standardization.

or the other pattern. The only languages that we are aware of as having been described as showing the same flexibility found in Sanskrit are Bantu languages like Chichewa. According to Alsina (1992), Chichewa freely admits both types of causative introduced above, and Alsina attributes this possibility to a number of other Bantu languages, Shona, Swahili and Kinyarwanda. However, Baker (1988: 161–167) discusses the Chichewa data and attributes the two causative structures to two distinct dialects; Baker (1988: 174–177) also discusses Kinyarwanda, but notes only the equivalent of the ACC-ACC causative.

Alsina (1996: 185–200) provides an analysis of two different causativization patterns in Romance, and argues that these correspond to the two types of causative found in Chichewa. Alsina treats dative-marked original subjects in Romance causatives as OBJ, contrasting with the alternative prepositional marking of the same argument, which he labels an OBL. However, Alsina does not properly distinguish OBJ from OBJ_θ, and allows two OBJ arguments in the same f-structure in violation of Consistency. As noted by Butt et al. (1997), the dative-marked original subjects in Romance causatives are better treated as OBJ_θ; thus Romance languages show only Baker's first type of causative.

In some languages, both patterns are found but with different sets of verbs. For example, in Marathi and some other modern Indo-Aryan languages, most verbs take the equivalent of the OBL-ACC causative, but a semantically identifiable subset of verbs, e.g. ingestive verbs, take the equivalent of ACC-ACC (Alsina and Joshi 1991). Çetinoğlu and Butt (2008) discuss data from Turkish which superficially appears to show both patterns at work with different sets of transitive verbs. Their analysis clearly demonstrates, however, that Turkish causatives of transitive verbs which apparently correspond to Baker's Rule 2 (and therefore to the Sanskrit ACC-ACC causative) in fact involve causativization of bivalent bases which take an OBJ_θ rather than OBJ alongside their subject. That is, the 'Rule 2' causatives in Turkish involve base verbs which are not transitive in the strictest sense of the word (i.e. in terms of taking a core OBJ argument). In terms of basic transitive verbs, then, Turkish consistently follows only Baker's Rule 1; the apparent Rule 2 causatives in fact follow the pattern for intransitive verbs. In contrast, in Sanskrit both patterns are found with the same verb (as illustrated above), and such verbs unambiguously involve OBJ rather than OBJ_θ, since the relevant argument becomes the subject in the (noncausative) passive:

(4) *kaṭaḥ* *kriyate*
mat.NOM make.PS.PRS.3SG
'A mat is made.'

Returning to the Sanskrit data, one approach to the difference between the ACC-ACC and OBL-ACC types is that the latter might be interpreted with a kind of 'passive' sense, e.g. (3) could be translated 'Yajñadatta caused the mat to be made by Devadatta'.[5] Yet the OBL-ACC causative is typologically well-paralleled, and

[5]So Kiparsky and Staal (1969: 102–103) argue that the OBL-ACC type is derived by applying

this type of causative is the only type of causative available in many languages, including modern Indo-Aryan languages, where a ‘passive’ interpretation would be untenable. In fact, the semantic difference between (2b) and (3) is not entirely clear. For Speyer (1886: 36–37), the distinction between the OBL-ACC and ACC-ACC causatives is to do with the intended expression: whether the causer acts on the original subject of base predicate (ACC-ACC) or on the original object (OBL-ACC). For Bubeník (1987), the difference between ACC-ACC and OBL-ACC causatives is contactive vs. non-contactive causation respectively. Following the native ancient grammarian Patañjali, Bubeník also refers to the degree of agency retained by the causee-agent (the original subject), stating that in the ACC-ACC type the original subject has less independence and agency, whereas in the OBL-ACC type the original subject has more independence and agency. On the other hand, Hock (1981) argues that at least in origin the OBL-ACC is marked, indicating lower agency and/or affectedness of the causee. Our investigations do not fully support any of these positions, and in this paper we remain agnostic about the difference between the OBL-ACC and ACC-ACC causatives, translating both neutrally as in (2b) and (3) above.

According to the indigenous grammarian Pāṇini, the default causative type is the OBL-ACC, and the ACC-ACC type is largely restricted to a subset of verbs of perception, consumption and making sound (*Aṣṭādhyāyī* 1.4.52–55). In reality, the ACC-ACC type is more widespread, and later grammarians extend its scope (see e.g. Joshi and Roodbergen 1975: 235–281). It remains true, however, that a subset of verbs, corresponding to Pāṇini's semantic classification, never show the OBL-ACC causative, such as the verb $\sqrt{}$*paṭh* ‘to recite’:[6]

(5) a. *māṇavako vedaṃ paṭhati*
boy.NOM veda.ACC recite.PRS.3SG
‘The boy recites the Veda.’

b. *Devadatto māṇavakaṃ vedaṃ pāṭhayati*
D.NOM boy.ACC veda.ACC recite.CAUS.PRS.3SG
‘Devadatta makes the boy recite the Veda.’

c. **Devadatto māṇavakena vedaṃ pāṭhayati*
D.NOM boy.INSTR veda.ACC recite.CAUS.PRS.3SG

passivization to the base predicate before causativization.

[6]It might be questioned whether it is justifiable to claim that some Sanskrit sentences are ungrammatical, as (5c), given that Sanskrit no longer has first-language speakers. Fortunately, the sophisticated and precise Sanskrit tradition of grammatical analysis delimited quite clearly what was and was not possible in Sanskrit; moreover, the surviving Sanskrit corpus is vast, running into the tens of millions of words at least, to the extent that the complete absence of a pattern is reasonable evidence for ungrammaticality, particularly when supported by native grammatical statements.

2.2 The passive

The finite passive in Sanskrit is formed from the root by means of mostly regular morphological processes: to the verbal root that usually stands in zero grade, the suffix *-ya-* and specialised mediopassive person/number endings are added to form the present passive. In other finite tenses (which are relatively uncommon in Classical Sanskrit), the passive is formally identical with the middle voice.[7] More common even than the finite present passive is the so-called 'past passive participle' in *-ta-/-na-*; in Vedic this 'participle' was only marginally integrated into the verbal paradigm (Lowe 2015b), but in Classical Sanskrit it is a full and productive part of the verbal paradigm, being the most common way of forming a past tense verb form. The past passive participle does function partially as a passive formation, but shows ergative alignment (the participle agreeing with the internal argument of a transitive verb, or the single argument of an intransitive verb), and is commonly used as a simple past tense, with no necessary passive interpretation.[8]

In the passive equivalent of an active clause, the object or core accusative argument of the verb is promoted to subject, and the demoted subject appears in the instrumental case. It is worth noting that there is no difference in basic (i.e. non-causative) passivization patterns between verbs which can (or always do) form OBL-ACC causatives and those that can only form ACC-ACC causatives. For example, the passives of both √*kṛ* (cf. 2–3) and √*paṭh* (5) work the same way:[9]

(6) a. *māṇavako vedaṃ paṭhati*
boy.NOM veda.ACC recite.PRS.3SG
'The boy recites the Veda.'

b. *vedaḥ paṭhyate māṇavakena*
veda.NOM recite.PASS.PRS.3SG boy.INSTR
'The Veda is recited by the boy.'

(7) a. *devadatto kaṭaṃ karoti*
D.NOM mat.ACC makes.PRS.3SG
'Devadatta makes a mat.'

b. *kaṭaḥ kriyate devadattena*
wood.NOM make.PASS.PRS.3SG D.INSTR
'A mat is made by Devadatta.'

This suggests that the internal arguments of all basic transitive verbs, both those like √*kṛ* and those like √*paṭh*, are fundamentally the same in argument structure

[7] Except in the aorist, which has a morphologically isolated passive formation restricted to the 3sg.

[8] For an account of the historical development of this participle, see Bynon (2005), with earlier references.

[9] The passive agent is optional, and often unexpressed; we do not represent it, or any other arguments, as optional in the glosses because in Sanskrit all arguments are omissible in context, even core object arguments.

terms, i.e. core object arguments in the active tenses which can be promoted to subjects in the passive.

2.3 Passive of causative

Passivization can also apply to causativized predicates.[10] As with the active causative, there are two possibilities. Either the first (i.e. the active subject), or the second (i.e. the active object), argument of the base predicate becomes the subject again in the passive causative. The first possibility is shown in (8), the second in (9).[11]

(8) *devadatto kaṭam kāryate yajñadattena*
D.NOM mat.ACC make.CAUS.PASS.PRS.3SG Y.INSTR

'Devadatta is made to make a mat by Yajñadatta.'

(9) *kaṭo devadattena kāryate yajñadattena*
mat.NOM D.INSTR make.CAUS.PASS.PRS.3SG Y.INSTR

'A mat is caused to be made by Devadatta by Yajñadatta.'

Although it is not the only conceivable possibility, it seems overwhelmingly likely that the promotion of the original subject of the base verb to subject represents the passive of the ACC-ACC causative, while the promotion of the original object to subject represents the passive of the OBL-ACC causative.[12]

This is supported by the fact that verbs which form only ACC-ACC causatives can also only passivize on the original subject. So, for the passive causative of √*paṭh*, only the type in (10) is found.

(10) *māṇavako vedaṃ pāṭhyate devadattena*
boy.NOM veda.ACC recite.CAUS.PASS.PRS.3SG D.INSTR

'The boy is made to recite the Veda by Devadatta.'

As passivization on the original object of an ACC-ACC causative is never found, this provides strong evidence that with such causatives, it is the original subject of the non-causative which is the core object argument in the causative, while the original object must be treated as a secondary object or oblique argument in the causative (since it cannot be passivized on).

[10] The passive can apply to causatives, but causativization cannot apply to passives. In the passive causative the passive suffix replaces the causative suffix, in all forms considered in this paper, so that the only morphological marking of the causative is the ablaut grade of the root.

[11] It is extremely rare for both instrumental arguments to be expressed in a passive like (9); cf. fn. 9.

[12] Setswana attests both types of passive causative, but has only one type of active causative, corresponding to the ACC-ACC causative of Sanskrit (R. Pretorius and A. Berg, p.c.); the same pattern is described for Kinyarwanda by Baker (1988: 174–180). In these Bantu languages the explanation is due to their being symmetrical object languages, meaning that the analysis will be somewhat different.

3 LFG analysis - preliminaries

The argument structure of causatives has been well studied within LFG, beginning with Alsina and Joshi (1991) and Alsina (1992, 1996), and developed most extensively by Butt (e.g. 1995, 1997, 1998, 2014).[13] In this paper, we follow Dalrymple et al. (2019) in integrating Butt's approach to complex predication with the model of argument structure developed by Kibort (2001, 2004, 2006, 2007). Kibort's argument structure model has been subject to precise formalization and integration with glue semantics by Findlay (2016), and this has been extended to a glue treatment of complex predication by Lowe (2015a, 2019). In the following we present 'traditional' argument structures modelled as complex semantic forms, but everything presented below could be unproblematically reformulated within a glue-based model.

As discussed by Lowe and Birahimani (2019), all the most important approaches to causative complex predicates and similar complex predication in LFG assume two types of predicate composition. Alsina and Joshi (1991) and Alsina (1992) first proposed the notion of *argument fusion*, whereby the causative predicate contains an argument position which is coindexed with an argument position of the embedded predicate. Alsina (1996) allows also a different kind of causative predicate, which does not involve coindexation of arguments; this is called *argument raising* by Butt (2014). For somewhat different reasons from Alsina, Butt (2014) also accepts both argument fusion and argument raising for the same complex predicates in Hindi-Urdu. We take the proposals of Butt (2014) to represent the most advanced and up-to-date treatment of the standard approach to predicate composition in LFG.

Lowe and Birahimani (2019) argue that the assumption of both argument fusion and argument raising for the same complex predicate is unsatisfactory for the analysis of morphological causatives in Siraiki, and for the same reasons we believe it to be equally unsatisfactory for causatives in Sanskrit. To begin with, Butt's (2014) approach requires that causative morphemes (or light verbs) are systematically ambiguous, showing both argument fusion and argument raising capabilities, and moreover, the choice of one or other possibility depends fundamentally on the embedded predicate. The assumption that both argument fusion and argument raising are possible cross-linguistically is entirely reasonable, and as argued by Butt (1998) the two notions respectively parallel the notions of raising and control in syntax. But we would not expect systematic ambiguity whereby single morphemes showed both possibilities, just as we do not find that raising verbs can systematically also be control verbs. In Butt's model, what is fundamentally a difference between the embedded predicate – whether it is transitive or intransitive – is modelled by means of an ambiguity in the embedding predicate. We feel that

[13] A brief analysis of the two causative patterns in Chichewa is provided by Bresnan et al. (2016: 341–343). However, they do not adopt a complex predicate approach to causativization, but assume monoclausal argument structures with variation in the a-structure classification of the agent. We will not consider this approach further here.

it would be preferable if the difference between causatives of transitive and intransitive verbs rather fell out naturally from the difference between the embedded predicates themselves.

The model of Butt (2014) is designed to deal with complex predicate patterns like causation specifically in languages like Hindi-Urdu, which show only the equivalent of the OBL-ACC causative to transitive verbs (i.e. they follow Baker's 1988 Causative rule 1). If we did try to transfer this to Sanskrit, where both OBL-ACC and ACC-ACC causatives are found, we would run into difficulties, as it would not be possible to license the ACC-ACC causatives without introducing a three-way ambiguity for the single causative morpheme: an argument fusion structure for the OBL-ACC causatives (with the causee specified as [+R]), an argument raising structure for causatives of intransitives, and a second argument fusion structure for the ACC-ACC causatives (with the causee specified as [−R]).[14]

Lowe and Birahimani (2019) show that it is possible, and argue that it is preferable, to model Siraiki causatives with reference to only one type of predicate composition, involving argument fusion, not raising.[15] This works out as long as the second argument of the causative predicate is unspecified for [±O/R] features, and given an independently required principle of ordering as to how argument linking proceeds. We show below that the Sanskrit data can be accounted for within the same kind of approach, although the details are more complicated.

Lowe and Birahimani (2019) propose an emendation to Kibort's (2007) universal valency template, augmenting it with the possibility of unspecified positions. For Sanskrit causatives, only a single unspecified position is required.[16] We therefore begin by assuming the valency template shown below:

(11) $$\left\langle \begin{array}{ccccccc} \text{arg}_1 & \text{arg}_2 & \text{arg}_3 & \text{arg}_4 & \text{arg}_5 & \ldots & \text{arg}_n \\ [-\text{O}/-\text{R}] & [-\text{R}] & [+\text{O}] & [\ ] & [-\text{O}] & & [-\text{O}] \end{array} \right\rangle$$

Below, we further argue that Sanskrit causatives and their passives require additional degrees of flexibility with regard to this valency template, raising questions over how far we can really consider this template universally fixed.

4 Analysis

The basic requirements for our analysis are the following. It is necessary to account for both the OBL-ACC and ACC-ACC causative formations, as well as the fact that with a certain subset of verbs, only the ACC-ACC type is possible. In addition, the argument structures for both types of causative must interact with the passive in the expected way.

[14]A conceivable alternative, argument raising applied to a transitive predicate, would not give a licit outcome for any transitive verb whose first argument is [−O].

[15]This does not mean that argument raising is not a possibility for causative structures in other languages; cf. the discussion above.

[16]This is because double causatives are not found; they first develop in early Middle Indic (Edgerton 1946).

In principle one might approach the difference between OBL-ACC and ACC-ACC causatives in two ways: either the causative morpheme itself is ambiguous, having two distinct argument frames or sets of properties which derive the required differences; or, there is a single causative argument frame associated with the (single) causative morpheme, and the two types of causative derive from some other point of variation. Following the principles discussed above and by Lowe and Birahimani (2019), we take the second course. For all (productive and semantically regular) causative formations in Sanskrit, then, we assume the argument frame below for the causative morpheme:

(12) $$\begin{array}{llccl} \textsc{caus} & \langle & \text{arg}_1, & \text{arg}_4 & \%\text{PRED} \quad \rangle \\ & & [-\text{o}] & [\ ] & \end{array}$$

We begin with the OBL-ACC causative. If we apply the causative predicate in (12) to an ordinary transitive verb like *kṛ* 'make', which appeared in (2), we obtain the following:

(13) $$\begin{array}{lcclccl} & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\ & | & | & & | & | & \\ \textsc{caus}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{`make'}\ \langle & \text{arg}_1 & \text{arg}_2 & \rangle\rangle \\ & [-\text{o}] & [\ ] & & ([-\text{o}]) & [-\text{r}] & \\ \hline & \text{SUBJ} & \text{OBL}_\theta & & & \text{OBJ} & \end{array}$$

Under Kibort's (2007) Mapping Principle, linking proceeds as follows: arg_1 links to the highest available grammatical function, then arg_2 links to the highest remaining grammatical function, and so on. For Siraiki causatives, Lowe and Birahimani (2019) argue that linking in Siraiki and other modern Indo-Aryan languages must likewise proceed according to arg index: arg_1 links first, then arg_2, etc., but crucially this is without consideration for embedding, so that the arg_2 of an embedded predicate will link before the arg_4 of the outer predicate.[17] In the case of (13), this obtains the right outcome: the arg_1 links to SUBJ, the embedded arg_2 links to OBJ, and then the arg_4 of the causative predicate links to OBL_θ.[18] However, since the embedded arg_2 is [−R], no other order of linking is possible: arg_1 maps to the highest available grammatical function, SUBJ, and the next highest available grammatical function is OBJ, which is also the only remaining [−R] function, and so must link to the embedded arg_2.

The passive of the OBL-ACC causative falls out unproblematically. Following Kibort (2007), we treat passivization in argument structure terms as the addition of the feature [+R] to the first argument of a predicate. Following the same principles of linking, the passive corresponding to (13) will therefore be:

[17]By 'embedded' here we refer to embedding in the argument structure; at f-structure the causatives are monoclausal. Note that in formulating her Mapping Principle, Kibort (2007) considers only simple argument structures without embedding.

[18]Note that when arguments are fused, the properties of the embedded argument which undergoes fusion are replaced by those of the fusing argument in the superordinate predicate.

(14)

	CAUSER	CAUSEE		AGENT	THEME
	|	|		|	|
CAUS ⟨	arg_1	arg_4	*kṛ* 'make' ⟨	arg_1	arg_2 ⟩⟩
	[−O]	[ ]		([−O])	[−R]
	[+R]				
	OBL$_\theta$	OBL$_\theta$			SUBJ

As it is specified with both [−O] and [+R], arg_1 must link to OBL$_\theta$. Arg_2 then links to the highest available GF, SUBJ, obtaining the desired passivization on the second argument (the original object) of the base predicate. On the traditional assumption that OBJ and OBL$_\theta$ are equally marked in terms of [±O/R] features, arg_4 can then link to OBL$_\theta$.[19] We assume that OBJ is not possible in this case, because then the OBJ, the core internal argument of the predicate, would scope over the SUBJ, the external argument, being related to the external event of causation while the SUBJ was related only to the embedded event.[20]

Turning now to the ACC-ACC causative, as illustrated in (2b), the first problem to address is the grammatical function of the second accusative argument, the second argument of the embedded predicate (*kaṭam* 'mat' in 2b). It is the first accusative argument, the causee-agent (*devadattam* 'Devadatta' in 2b), which must be the core object in the ACC-ACC causative, not least because it becomes the subject in the passive of this causative (8). The second accusative argument cannot therefore be an OBJ. It could be either OBL$_\theta$ or OBJ$_\theta$, but neither of these is compatible with the [−R] specification of the arg_2 argument slot.[21]

To resolve this difficulty, we propose that instead of an arg_2 specified as [−R], ordinary transitive verbs in Sanskrit have an underspecified second argument slot which in Kibort's terms is either arg_2 or arg_3, i.e. $\mathrm{arg}_{2/3}$. This argument slot then either has the specification [−R] or the specification [+O]. There is cross-linguistic support for such a featural specification: [−R]/[+O] is exactly what Alsina and Mchombo (1993) proposed for internal arguments (except beneficiaries and recipients) to account for applicatives in Bantu.[22]

In the unmarked case, where a verb like *kṛ* 'make' is used in a noncausative, active, sentence (as in 2a), the difference is moot: either will map to OBJ. In the active causative, the difference is also moot in exactly the same way. So, we could rewrite (13) above with the new underspecified value for the second argument of 'make', and there will be no change in the linking:

[19] But note that for Her (2013), OBJ should be less marked than OBL$_\theta$ because [+O] is less marked than [+R].

[20] This could be considered a general constraint on complex argument structures; it is certainly hard to see how such a structure could make sense, or why such a structure would ever be desired.

[21] See Lowe (2017: 33–34) for accusative case OBL$_\theta$ arguments in Sanskrit.

[22] This proposed specification for internal arguments was accepted by Bresnan and Moshi (1990).

(15)

	CAUSER	CAUSEE		AGENT	THEME	
CAUS ⟨	arg$_1$	arg$_4$	*kṛ* 'make' ⟨	arg$_1$	arg$_{2/3}$	⟩⟩
	[−O]	[ ]		([−O])	[−R]/[+O]	
	SUBJ	OBL$_\theta$			OBJ	

In the passive of the OBL-ACC, on the other hand, the [−R] feature must apply to the underspecified argument, since otherwise the predicate would lack a subject and violate the Subject Condition (Bresnan and Kanerva 1989, Berman 1999, 2003).[23]

(16)

	CAUSER	CAUSEE		AGENT	THEME	
CAUS ⟨	arg$_1$	arg$_4$	*kṛ* 'make' ⟨	arg$_1$	arg$_{2/3}$	⟩⟩
	[−O]	[ ]		([−O])	[−R]/[+O]	
	[+R]					
	OBL$_\theta$	OBL$_\theta$			SUBJ	

How does this help us with the ACC-ACC causative? As discussed above, the difference between these two causative argument structures should derive from some point of variation other than assuming two different versions of the causative predicate itself. Above, we derived the OBL-ACC causative by following Lowe and Birahimani (2019) in interpreting Kibort's Mapping Principle such that argument linking proceeds according to arg index, regardless of embedding. But there is one other logically possible way of interpreting Kibort's Mapping Principle with respect to complex predicates. That is, linking may proceed linearly, 'left-to-right'; or to put it another way, it may proceed according to arg index but with regard to embedding (so that an arg$_4$ which is in a superordinate predicate links before an arg$_2$, etc.).[24] If we assume that both are possible in Sanskrit, we immediately get the ACC-ACC causative:

(17)

	CAUSER	CAUSEE		AGENT	THEME	
CAUS ⟨	arg$_1$	arg$_4$	*kṛ* 'make' ⟨	arg$_1$	arg$_{2/3}$	⟩⟩
	[−O]	[ ]		([−O])	[−R]/[+O]	
	SUBJ	OBJ			OBJ$_\theta$	

The arg$_1$ of the causative predicate links to SUBJ, the arg$_4$ of the causative predicate links to OBJ, and then the arg$_{2/3}$ of the embedded predicate must link to

[23] The same applies to the noncausative passive.

[24] Kibort (2007) considers only simple structures with no embedding, i.e. structures in which linking according to arg index and linking according to left-to-right position are indistinguishable in their effects. It is a neat outcome of our analysis that the two possible interpretations of Kibort's Mapping Principle license the difference between the two causative structures under discussion.

OBJ$_\theta$ by virtue of the [+O] specification, since both [−R] arguments are already linked.

The passive of the ACC-ACC causative also falls out directly. With left-to-right linking, the arg_4 of the causative predicate will link to SUBJ, and the embedded $\text{arg}_{2/3}$ to OBJ:

(18)

$$
\begin{array}{lcccccl}
 & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\
 & | & | & & | & | & \\
\text{CAUS}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{`make'}\ \langle & \text{arg}_1 & \text{arg}_{2/3} & \rangle\rangle \\
 & [-\text{O}] & [\,] & & ([-\text{O}]) & [-\text{R}]/[+\text{O}] & \\
 & [+\text{R}] & & & & & \\
\hline
 & \text{OBL}_\theta & \text{SUBJ} & & & \text{OBJ} &
\end{array}
$$

Both types of causative, and their passives, thus derive from a single causative argument frame, a frame which will also account for causatives of intransitives. Recall that under Butt's (2014) approach, three distinct argument frames would be required for the single causative morpheme to account for these patterns.

It remains to account for the set of transitive verbs which can only form ACC-ACC causatives, such as *paṭh* 'recite' illustrated in (5). These verbs do not otherwise differ from the majority of transitive verbs in Sanskrit, and although there are semantic patterns, there are no absolute criteria by which this subset may be distinguished semantically.[25]

Perhaps the simplest solution would be to state an informal constraint to the effect that a certain subset of verbs require linking to proceed left-to-right and disallow linking according to arg index. Another way of looking at this would be to say that the verbs subject to this constraint require their arg_1s to link before their arg_2s, even when embedded under a causative predicate; and this perspective might be explicable e.g. in terms of the relative affectedness of the respective arguments.

Such a constraint cannot easily be formalized, however. A formal solution is possible, within the framework of our analysis and the assumptions made so far, but it requires us to assume further degrees of flexibility in Kibort's (2007) valency template. Beyond arg_3, Kibort assumes a potentially unrestricted number of arg slots, all with the specification [−O]; these are slots which link to various kinds of oblique arguments. But we could assume that beyond the initial core of three or four arg slots, there is more flexibility. Thus it might be possible, for example, to have an arg slot, say arg_6, with a specification [−R]/[+O]. This is the same specification as the $\text{arg}_{2/3}$ assumed above for ordinary transitive verbs in Sanskrit.

[25] Alsina (1996: 196–197) notes a partial parallel in Romance languages, where verbs with experiencer subjects can only form the causative with original subject expressed as a dative secondary object, and cannot form the alternative causative with original subject as oblique. Alsina argues that the restriction is a restriction on thematic roles: in Romance, at least, an oblique causee may not be an experiencer. The same explanation cannot apply to Sanskrit, since some verbs subject to this constraint take agent subjects. In Sanskrit the difference appears more to do with the affectedness of the object; we do not pursue the details of this further here, but note that our account assumes a difference between the objects of these verbs and of verbs not subject to the constraint.

But crucially the arg index of this slot is higher than the index of the arg_4 of the causative predicate.

We could therefore assume that the two sets of transitive verb in Sanskrit differ in no way (syntactically, at least), and that causativization applies in exactly the same way for both sets. But whereas with the *kṛ* 'make' type above, the flexibility in order of arg linking gave two possibilities, one in which the embedded $\text{arg}_{2/3}$ linked before the causative's arg_4, and one in which the converse order applied, with the *paṭh* 'recite' type both orders of arg linking give the same result: the causative arg_4 cannot but link before the arg_6 of the embedded predicate, because it both precedes it in 'left-to-right' order, and has a lower index.

(19)

	CAUSER	CAUSEE		AGENT	THEME	
	|	|		|	|	
CAUS ⟨	arg_1	arg_4	*paṭh* 'recite' ⟨	arg_1	arg_6	⟩⟩
	[−O]	[ ]		([−O])	[−R]/[+O]	
	SUBJ	OBJ			OBJ_θ	

(20)

	CAUSER	CAUSEE		AGENT	THEME	
	|	|		|	|	
CAUS ⟨	arg_1	arg_4	*paṭh* 'recite' ⟨	arg_1	arg_6	⟩⟩
	[−O]	[ ]		([−O])	[−R]/[+O]	
	[+R]					
	OBL_θ	SUBJ			OBJ	

In this section we have explored how we can model Sanskrit causatives and passive causatives within the framework of Kibort (2007), with the augmentations of Lowe and Birahimani (2019). Taking the two possible interpretations of Kibort's Mapping Principle when applied to argument structures with embedding, we have been able to neatly explain the alternation between the OBL-ACC and ACC-ACC causatives. Our analysis requires us to assume that core internal arguments are specified as [−R]/[+O], in line with proposals made for Bantu languages.

The assumption of an underspecified [−R]/[+O] requires more flexibility in Kibort's 'universal' argument structure template, on top of the assumption of fully unspecified slots. If we then accept also the idea that we might have slots such as arg_6 with a specification [−R]/[+O], as suggested immediately above, we move towards a point where our approach requires such flexibility in the template that we can no longer really call it universal. The central innovations of Kibort's model – the use of indexed arg positions, separating syntactic valency from semantic roles, and the unified Mapping Principle which depends on the arg indices – remain. But we would then essentially have reached the point where a strict and universal association between arg positions and [±O/R] features is no longer possible. In itself this is undesirable, as it makes the model less constrained and more stipulatory. On the other hand, there are also some attractive elements to our analysis, including the fact that it is able to model the Sanskrit data, in particular the coexistence of OBL-ACC and ACC-ACC causatives, much more satisfactorily than previous proposals. We do not claim that the proposals we have made here are perfect, or that they should be the final word, but we present them as an exploration of what can be done, and what must be done, in order to effectively model the argument structure of Sanskrit causatives within the 'Kibortian' framework adopted here.

5 Exploring an alternative

The analysis presented in the preceding section works: it accounts for the data, has some attractive aspects, and does not suffer from the weaknesses of Butt's (2014) model discussed above. At the same time, it requires certain assumptions or stipulations, and requires us to relax the strict argument structure template of Kibort (2007) to the point where we must question its universality. In this section we explore an alternative approach (though still presented within the framework of Kibort 2007: for consistency).

As noted above, the data analysed in this paper is parallel to that for Chichewa as analysed by Alsina (1992), and similar also to the analysis of Romance in Alsina (1996). Alsina's analysis is rather different from our own, and in this section we explore whether it might provide a better account of the Sanskrit data than the approach developed above.

The main distinguishing feature of Alsina's approach is the assumption that argument fusion may be with potentially any argument of the embedded predicate. Alsina (1992, 1996) permits the causee argument of the causative predicate to fuse either with the first or second argument of an embedded transitive predicate. If it fuses with the first argument, the equivalent of the ACC-ACC causative results; if it fuses with the second argument, the equivalent of the OBL-ACC causative results. Note that, like our analysis in the previous section, Alsina's analysis assumes a single invariant causative argument frame; the variation comes in the fusion of arguments (whereas the variation in our account above comes in the order of arg linking).

Let us see how this works if we update Alsina's proposals and attempt to integrate them into the Kibortian argument structure approach adopted above. The analysis of the ACC-ACC causative (21), and its passive (22), is almost identical to what we presented above:

(21)
$$\begin{array}{lcccccc}
 & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\
 & | & | & & | & | & \\
\text{CAUS}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{`make'}\ \langle & \text{arg}_1 & \text{arg}_{2/3} & \rangle\rangle \\
 & [-\text{o}] & [\ ] & & ([-\text{o}]) & [-\text{r}]/[+\text{o}] & \\
\hline
 & \text{SUBJ} & \text{OBJ} & & & \text{OBJ}_\theta &
\end{array}$$

$$
\begin{array}{lcccccc}
(22) & & & & & & \\
 & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\
 & | & | & & | & | & \\
\text{CAUS}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{‘make’}\ \langle & \text{arg}_1 & \text{arg}_{2/3} & \rangle\rangle \\
 & [-\text{O}] & [\ ] & & ([-\text{O}]) & [-\text{R}]/[+\text{O}] & \\
 & [+\text{R}] & & & & & \\
\hline
 & \text{OBL}_\theta & \text{SUBJ} & & & \text{OBJ} &
\end{array}
$$

In order to get the linking right in (21), we need to assume left-to-right linking to ensure that the arg_4 links to OBJ rather than the embedded $\text{arg}_{2/3}$. We also still require the embedded second argument to be $\text{arg}_{2/3}$, i.e. with the features [−R]/[+O], since this argument can surface as SUBJ (in the noncausative passive), OBJ or OBJ_θ.

For the OBL-ACC causative, we require argument fusion to take place between the second argument of the causative predicate and the second argument of the embedded predicate:

$$
\begin{array}{lcccccc}
(23) & & & & & & \\
 & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\
 & | & | & & | & | & \\
\text{CAUS}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{‘make’}\ \langle & \text{arg}_1 & \text{arg}_{2/3} & \rangle\rangle \\
 & [-\text{O}] & [\ ] & & [-\text{O}] & ([-\text{R}]/[+\text{O}]) & \\
\hline
 & \text{SUBJ} & \text{OBJ} & & \text{OBL}_\theta & &
\end{array}
$$

$$
\begin{array}{lcccccc}
(24) & & & & & & \\
 & \text{CAUSER} & \text{CAUSEE} & & \text{AGENT} & \text{THEME} & \\
 & | & | & & | & | & \\
\text{CAUS}\ \langle & \text{arg}_1 & \text{arg}_4 & \textit{kṛ}\ \text{‘make’}\ \langle & \text{arg}_1 & \text{arg}_{2/3} & \rangle\rangle \\
 & [-\text{O}] & [\ ] & & [-\text{O}] & ([-\text{R}]/[+\text{O}]) & \\
 & [+\text{R}] & & & & & \\
\hline
 & \text{OBL}_\theta & \text{SUBJ} & & \text{OBL}_\theta & &
\end{array}
$$

We again need to assume that linking takes place left-to-right, to ensure that in both active and passive the arg_4 of the causative predicate links before the arg_1 of the embedded predicate. Everything else follows without difficulty.

One issue with this approach (as it was with our approach above) is how to prevent the OBL-ACC causative with verbs like $\sqrt{\textit{paṭh}}$ ‘to recite’. In the approach presented in the previous section, we proposed a formal solution on the basis that these verbs have a higher index for their second arg slot, thereby neutralizing the difference between the two ways of ordering linking. In the approach currently under discussion, however, the variation comes not from the order of linking arguments, but from the possibility of argument fusion with either the first or second argument of the embedded predicate. It is not immediately obvious to us how to formally prevent argument fusion with the second argument of only certain verbs, but presumably an informal constraint could be formulated to the relevant effect.

Butt (1998) argues strongly against Alsina’s freedom in licensing argument fusion, objecting that permitting merger with any argument is insufficiently constrained. Butt (1998) proposes that argument fusion is the argument structure correspondent of control, and argues for the constraint that fusion can only ever be with the first argument of the embedded predicate ("Restriction on Argument Fusion"). Within the Kibortian model adopted here, there are further arguments against fusion with the second argument. Fusion with the first argument is always monotonic in the sense that no arg slots are lost: the arg_1 of the embedded predicate does not link, but there is still an arg_1, introduced by the causative predicate. This is crucial to Lowe's (2015a) glue-based formalization of complex predicates: the arg_1 label of the lexical predicate becomes the arg_1 of the causative predicate, and the properties of the embedded arg_1 are transferred to the arg_4 of the causative predicate, but no arg positions are lost. If we permit fusion with the second argument of the embedded predicate, we end up with two arg_1s, and an $\text{arg}_{2/3}$ which has to essentially disappear. Within the monotonic glue formalization of Lowe (2015a), it would be impossible to deal with this.

Furthermore, at least in Sanskrit, we do not find a third pattern with causatives of ditransitive verbs which could be construed as argument fusion with the indirect object.[26] If it were possible for argument fusion to target the second argument of an embedded predicate, why should it not also be able to target a third argument, when present?

Overall, both Alsina's approach, as presented here in updated form, and the approach proposed above share some common features, and both require assumptions over and above those found in previous literature. We concur with Butt (1998) in restricting argument fusion to targeting the highest argument of the embedded predicate, and thus prefer the approach presented in §4. At the same time, we acknowledge that that approach takes us quite far from the original intention of Kibort's universal argument structure template.

6 Conclusion

In this paper, we have presented data of causatives and passive causatives in Sanskrit, and have explored possible LFG analyses broadly within the argument structure model of Kibort (2007) and the approach to complex predicates of Butt (2014).

Causatives and their passives in Sanskrit are of interest because the two cross-linguistically common patterns of causativization are found together in the same language, and even with the same verb forms. Under our proposed analysis, the variation between the ACC-ACC and OBL-ACC causatives derives from two possibilities in the order of linking argument slots to grammatical functions. Linking according to arg index without respect for embedding, as proposed for modern Indo-Aryan causatives by Lowe and Birahimani (2019), gets the OBL-ACC causative, and its corresponding passive. Linking left-to-right gets the ACC-ACC causative.

[26]Due to lack of space we have not discussed causatives of ditransitives above. Essentially, they work just like causatives of transitives, but with an additional beneficiary argument, which appears in the dative or genitive case, and this does not change in causative structures.

Our proposed analysis is not perfect, but is superior to existing LFG analyses in some respects. It requires increased flexibility in the Kibortian approach to argument structure, but depends on some of its key innovations, in particular the notion of indexed arg slots. We believe that it will be profitable to further explore the possibilities and limits of this approach to argument structure, and to complex predicates, in order to better understand its advantages and disadvantages relative to earlier models of linking in LFG.

References

AISSEN, JUDITH L. (1979). *The syntax of causative constructions*. New York: Garland.

ALSINA, ALEX (1992). 'On the argument structure of causatives'. *Linguistic Inquiry 23*(4), pp. 517–555.

ALSINA, ALEX (1996). *The Role of Argument Structure in Grammar*. Stanford, CA: CSLI Publications.

ALSINA, ALEX and SMITA JOSHI (1991). 'Parameters in Causative Constructions'. In L. Dobrin, L. Nichols, and R. M. Rodriguez (eds.), *Proceedings from the 27th Regional Meeting of the Chicago Linguistic Society (CLS)*, Chicago, IL: Chicago Linguistic Society, pp. 1–16.

ALSINA, ALEX and SAM A. MCHOMBO (1993). 'Object asymmetries and the Chichewa applicative construction'. In Sam A. Mchombo (ed.), *Theoretical Aspects of Bantu Grammar*, Stanford, CA: CSLI Publications, pp. 17–45.

BAKER, MARK C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

BERMAN, JUDITH (1999). 'Does German Satisfy the Subject Condition?' In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG99 Conference*, Stanford, CA: CSLI Publications.

BERMAN, JUDITH (2003). *Clausal Syntax of German*. Stanford, CA: CSLI Publications.

BLOCH, JULES (1965). *Indo-Aryan from the Vedas to Modern Times*. Paris: Adrien-Maisonneuve. Translated by Alfred Master.

BRESNAN, JOAN, ASH ASUDEH, IDA TOIVONEN, and STEPHEN WECHSLER (2016). *Lexical-Functional Syntax*. Oxford: Wiley-Blackwell. Second edition. First edition by Joan Bresnan, 2001, Blackwell.

BRESNAN, JOAN and JONNI M. KANERVA (1989). 'Locative Inversion in Chicheŵa: A Case Study of Factorization in Grammar'. *Linguistic Inquiry 20*, pp. 1–50.

BRESNAN, JOAN and LIOBA MOSHI (1990). 'Object Asymmetries in Comparative Bantu Syntax'. *Linguistic Inquiry 21*(2), pp. 147–185.

BUBENÍK, VÍT (1987). 'Passivized causatives in Sanskrit and Prākrits'. *Linguistics 25*, pp. 687–704.

BUTT, MIRIAM (1995). *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI Publications.

BUTT, MIRIAM (1997). 'Complex Predicates in Urdu'. In Alex Alsina, Joan Bresnan, and Peter Sells (eds.), *Complex Predicates*, Stanford, CA: CSLI Publications.

BUTT, MIRIAM (1998). 'Constraining Argument Merger Through Aspect'. In Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa (eds.), *Complex Predicates in Nonderivational Syntax*, San Diego, CA: Academic Press, pp. 73–113.

BUTT, MIRIAM (2014). 'Control vs. Complex Predication: Identifying Non-Finite Complements'. *Natural Language & Linguistic Theory 32*(1), pp. 165–190.

BUTT, MIRIAM, MARY DALRYMPLE, and ANNETTE FRANK (1997). 'An architecture for linking theory in LFG'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

BYNON, THEODORA (2005). 'Evidential, Raised Possessor and the Historical Source of the Ergative Construction in Indo-Iranian'. *Transactions of the Philological Society 103*(1), pp. 1–72.

CARDONA, GEORGE (1978). 'Relations between causatives and passives in Indo-Iranian'. *Studies in the Linguistic Sciences 8*(2), pp. 1–42.

ÇETINOĞLU, ÖZLEM and MIRIAM BUTT (2008). 'Turkish Non-Canonical Objects'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, Stanford, CA: CSLI Publications, pp. 214–234.

DALRYMPLE, MARY, JOHN J. LOWE, and LOUISE MYCOCK (2019). *The Oxford reference guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

DESHPANDE, MADHAV M. (1991). 'Ditransitive passive in Pāṇini'. *Indo-Iranian Journal 34*, pp. 19–35.

EDGERTON, FRANKLIN (1946). 'Indic causatives in *-āpayati* (*-āpeti*, *-āvei*)'. *Language 22*(2), pp. 94–101.

FINDLAY, JAMIE Y. (2016). 'Mapping Theory without Argument Structure'. *Journal of Language Modelling 4*(2), pp. 293–338.

HER, ONE-SOON (2013). 'Lexical Mapping Theory Revisited'. In Tracy Holloway King and Valeria de Paiva (eds.), *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, Stanford, CA: CSLI Publications, pp. 47–59.

HOCK, HANS HENRICH (1981). 'Sanskrit causative syntax: A diachronic study'. *Studies in the Linguistic Sciences 11*(2), pp. 9–33.

JAMISON, STEPHANIE W. (1983). *Function and Form in the* -áya-*Formations of the Rig Veda and Atharva Veda*. Göttingen: Vandenhoeck & Ruprecht.

JOSHI, S. D. and J. A. F. ROODBERGEN (1975). *Patañjali's Vyākaraṇa-mahābhāṣya: Kārakāhnika (P. 1.4.23–1.4.55)*. Poona: University of Poona.

KIBORT, ANNA (2001). 'The Polish passive and impersonal in Lexical Mapping Theory'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, Stanford, CA: CSLI Publications, pp. 163–183.

KIBORT, ANNA (2004). *Passive and Passive-Like Constructions in English and Polish*. Ph.D. thesis, University of Cambridge.

KIBORT, ANNA (2006). 'On Three Different Types of Subjectlessness and How to Model Them in LFG'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, Stanford, CA: CSLI Publications.

KIBORT, ANNA (2007). 'Extending the Applicability of Lexical Mapping Theory'. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, Stanford, CA: CSLI Publications, pp. 250–270.

KIPARSKY, PAUL and J. F. STAAL (1969). 'Syntactic and Semantic Relations in Pāṇini'. *Foundations of Language 5*(1), pp. 83–117.

KULIKOV, LEONID (2012). *The Vedic* -ya-*presents: Passives and intransitivity in Old Indo-Aryan*. Amsterdam: Rodopi.

KULIKOV, LEONID (2013). 'Constraints on the causative derivation in early Vedic: Evidence for a diachronic typology of transitivity'. *Poznań Studies in Contemporary Linguistics 49*(1), pp. 79–101.

LOWE, JOHN J. (2015a). 'Complex predicates: an LFG+Glue analysis'. *Journal of Language Modelling 3*(2), pp. 413–462.

LOWE, JOHN J. (2015b). *Participles in Rigvedic Sanskrit: The Syntax and Semantics of Adjectival Verb Forms*. Oxford: Oxford University Press.

LOWE, JOHN J. (2017). *Transitive Nouns and Adjectives: Evidence from Early Indo-Aryan*. Oxford: Oxford University Press.

LOWE, JOHN J. (2019). 'Argument alternations in complex predicates: an LFG+glue perspective'. *Linguistic issues in Language Technology 17*(2), pp. 1–16.

LOWE, JOHN J. and ALI H. BIRAHIMANI (2019). 'The argument structure of Siraiki causatives'. In Miriam Butt, Tracy Holloway King, and Ida Toivonen (eds.), *Proceedings of the LFG19 Conference*, Stanford, CA: CSLI.

MASICA, COLIN P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

RENOU, LOUIS (1961). *Grammaire sanscrite*. Paris: Adrien-Maisonneuve. Second edition: revised, corrected and augmented.

SPEYER, JACOB SAMUEL (1886). *Sanskrit Syntax*. Leiden: Brill.

An LFG Analysis of Setswana Auxiliary Verb Phrases Indicating Tense

Rigardt Pretorius
North-West University, South Africa

Ansu Berg
North-West University, South Africa

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 233–250

http://csli-publications.stanford.edu/LFG/2019

Keywords: auxiliary verbs, Setswana, tense, f-structure

Pretorius, Rigardt, & Berg, Ansu. 2019. An LFG Analysis of Setswana Auxiliary Verb Phrases Indicating Tense. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 233–250. Stanford, CA: CSLI Publications.

Abstract

This paper focusses on the treatment of the semantic values of Setswana auxiliary verbs in cases where more than one auxiliary verb appears in the same auxiliary verb phrase (VPAux). The aim is to give an overview of the treatment of the tense/aspect features of these auxiliary verb phrases at f-structure in LFG. The minimal structure of a VPAux consists of an auxiliary verb followed by an obligatory phrase that may be a verb phrase that includes a main verb (VPMain) or a copulative verb (VPIdcop, VPDescop, VPAsscop) or the obligatory phrase may be another VPAux. We give an outline of the categories expressed in main verbs followed by a brief overview of the characteristics of Setswana auxiliary verbs. We then propose a treatment of the tense and aspect features of auxiliary verbs within the VPAux.

1. Contextualisation

Setswana, a Bantu language in the Sotho language group (group S31) (Maho, 2003:639-651), is one of the official languages of South Africa. As a result of the agglutinative nature of Bantu languages, verbs in these languages are inflected with numerous morphemes. Verbs can therefore be morphologically complex in Setswana. Lexically, Setswana verbs are classified into main (independent) verbs, copulative verbs and auxiliary verbs based on their morphological, syntactic and semantic features. The morphosyntactic features of Setswana verbs have been treated in grammars such as Cole (1955), Krüger (2006, 2013a, 2013b), Ranamane (2009) and Berg (2018) among others, while Brits *et al.* (2005) and Pretorius *et al.* (2009) amongst others address the specific computational challenges that these typological features present.

Auxiliary verbs always occur with main or copulative verbs as complements In (1)[1] the auxiliary verb **o ne** is followed by the main verb **wa ya** (you then go) and in (2) the auxiliary verb **ke ne** is followed by the main verb **ka ya** (I then go). Auxiliary verb phrases in Setswana cannot be abbreviated as in the case of English where ellipsis may appear resulting in the elision of the

[1] **Abbreviations and morphological tags:**

AgrObj	object agreement morpheme	P1	first person
AgrSubj	subject agreement morpheme	PassSuf	passive suffix
AgrSubjCons	consecutive subject agreement morpheme	PerfSuf	perfect suffix
ApplSuf	applicative suffix	pl	Plural
Aux	auxiliary verb	PotPre	potential morpheme
CausSuf	causative suffix	PresPre	present tense morpheme
DevSuf	deverbative suffix	ProgPre	progressive prefix
FutPre	future prefix	RecSuf	reciprocal suffix
ImpSuf	imperative suffix	ReflPre	reflexive morpheme
Interj	Interjection	RelSuf	relative suffix
IntPart	interrogative particle	sg	Singular
LocPart	locative particle	TempPart	temporal particle
NegPre	negative morpheme,	VEnd	verbal ending
Npre	noun prefix	1 2 3 ...	noun classes 1 – 20

complement of the auxiliary verb. The auxiliary verb **ke ne** in (3) is an invalid structure as it is compulsory for an auxiliary verb to be followed by a complement.

(1) A o ne wa ya kwa toropong?

a	*o-ne*	*wa-y-a*	*kwa*
IntPart	AgrSubjP2sg-Aux	AgrSubjConsP2sg-go-VEnd	LocPart

(ne-)-toropo-ing
NPre9-town-LocSuf
Did you go to town?

(2) Ee, ke ne ka ya.

ee	*ke-ne*	*ka-y-a*
Interj	AgrSubjP1sg-Aux	AgrSubjConsP1sg-go-VEnd

Yes, I did go.

(3) *Eê, ke ne…

ee	*ke-ne*
Interj	AgrSubjP1sg-Aux

*Yes, I …

The features of auxiliary verbs thus have to be considered with those of their complements (Setshedi, 1974; Pretorius, 1997). In the discussion of auxiliary verb phrases, the treatment of their semantic values seems to be illusive. The focus in this paper is on auxiliary verb phrases in Setswana and more specifically on the treatment of their semantic values in cases where more than one auxiliary verb appears in the same phrase. This will be done by giving a brief overview of the characteristics of Setswana auxiliary verbs followed by an outline of the categories expressed in main verbs. Auxiliary verbs with copulative complements are not discussed in this paper. The categories of tense and aspect are always present in verbs even if they are not overtly marked in the morphology (Cole, 1955:235-285; Pretorius, 1997:344-347). As tense/aspect (T/A) are the prominent connection between the grammatical categories of the verb and its use with auxiliary verbs, the focus will be on T/A. In the discussion of the auxiliary verb phrases the focus will be on instances where more than one auxiliary verb is present (juxtaposed) in order to give a literal representation of word order and constituency in c-structure. The objective is the treatment of the T/A values of these auxiliary verb phrases at f-structure in LFG. We expand on the recently developed LFG grammar for Setswana (Berg 2018). We use data from mother tongue speakers as well as grammars (Cole, 1955; Krüger, 2006; Ranamane, 2009) for analysis and discussion.

2. Setswana main verbs

2.1 Categories expressed by main verbs

Setswana main verbs typically express the grammatical features of tense, aspect, mood and polarity (TAMP). The main verb may include inflectional prefixes as well as derivational and inflectional suffixes and these affixes contribute various semantic values. In terms of morphotactics, each affix occupies a specific slot in the morphological structure of a main verb (Krüger, 2006:268):

NegPre (**ga**) + AgrSubj + NegPre (**sa**, **se**, **a**) + PresPre/ProgPre/PotPre + FutPre + AgrObj/ReflPre + Root +
Productive suffixes + VEnd + RelSuf/ImpSuf

Specific morphological information pertaining to the prefixes (negative morpheme **ga**, subject agreement morphemes, negative morphemes **sa**, **se** and **a**, present tense morpheme, progressive morpheme, future tense morpheme, potential morpheme, object agreement morphemes and reflexive morpheme) inflect essential information. For example, the negative prefixes, subject agreement morphemes, present tense morpheme, progressive morpheme and future tense morpheme contribute to tense, aspect, mood and polarity (TAMP). The productive suffixes include the causative, applicative, reciprocal and the passive extensions as well as the perfect suffix. The inflectional perfect morpheme specifies either a past tense or perfective aspect. The verbal endings are inflectional morphemes and their form (**a**, **e**, **ê**) is determined by mood, tense and polarity information. A relative suffix or imperative suffix may be added after the verbal ending.

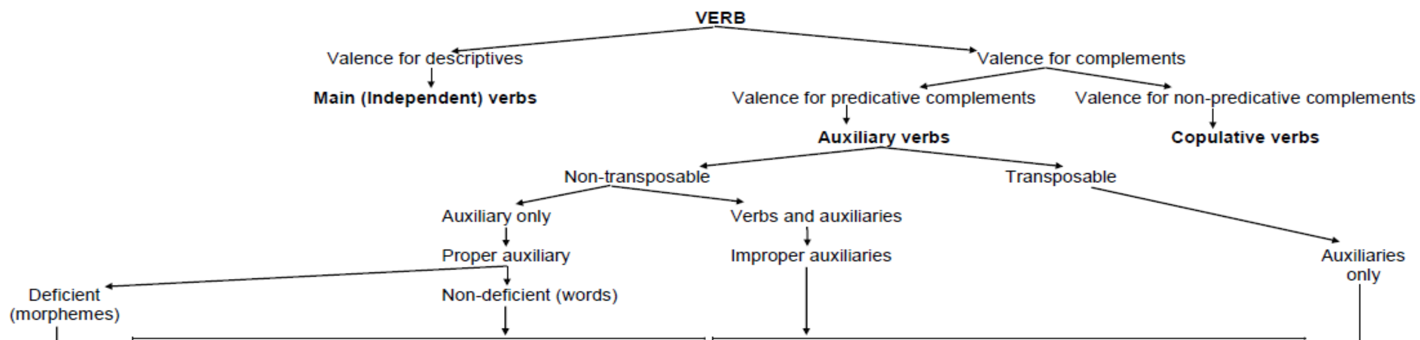
3. Setswana auxiliary verbs

The origin, meaning and function of auxiliary verbs in Setswana show that they have a metaphorical basis. Diachronically auxiliaries come from main verbs with lexical meanings. This lexical meaning, even though it is bleached, informs the meaning of the auxiliary. Many auxiliary verbs still show semantic affinity with main verbs. This affinity is morphologically and syntactically related to the sub-classes identified in the classification of the auxiliary verbs.

3.1 Classification of Setswana auxiliary verbs

As presented in Diagram 1 (Pretorius 1997:89), the verb in Setswana has three sub-categories, viz. auxiliary verbs, main (independent) verbs and copulative verbs (Pretorius, 1997:14, 76-81).

Auxiliary verbs are classified into a non-transposable category and a transposable category which semantically has a conjunctive function. In the non-transposable category, a distinction is made between auxiliary verbs that can only function as auxiliaries (proper auxiliaries) and auxiliary verbs that can



	Used in negative structures	Semantic value of logical time	Coordinative consecutive	Inferential/ stative	Relative past/ perfective	Semantic value of duration/ continuation	Semantic value of logical time	Semantic value of successive continuation	Semantic value of simultaneous execution	Semantic value of repetition	Conjunctive function
Continuative	-bisa	-bolo	-ba	-bo	-ne	-aga	-akofa	-fela	-nama	-atisa	e bile
-a-	-ise	-sena	(-bo)			-nna	-dika	-tloga	-sala	-boa	e kete
Progressive	-ka	-jafile	-bile			-ntse	- letse	-feta	-eta/nna	(/-boela)	e re
-sa	-ketla	-kile				-tlhola	-lala	-tsamaya			gobo
Futuritive	-ne	-ile					-phakela	-tla			gonne
-tla	-nke	-sale					-tsoga	-fitlha			gobane
Potential		-ekile									gore
-ka		-setse									go tla
											go tloga
											go fitlhela
											go tsamaya
											ntswa/etswa
											mme

Diagram 1: Classification of Setswana auxiliary verbs (Pretorius, 1997:89)

function as verbs or auxiliaries (improper auxiliaries[2]). Within the categories of proper and improper auxiliaries, constituents are classified into groups on semantic grounds. The proper auxiliaries function as morphemes (deficient) or words (non-deficient). The deficient auxiliaries are morphemes and they may be included in the morphology of main and auxiliary verbs. Pretorius (2004:198) indicates that as a result of the development of auxiliaries, some auxiliaries have lost their status as words and have rank shifted to prefixed verbal morphemes. The auxiliary verbs with conjunctive function are not treated here. The morphological, syntactic and semantic characteristics of auxiliary verbs are introduced following Diagram 1.

3.2 Characteristics of auxiliary verbs

3.2.1 Morphological characteristics

The morphological structure of an auxiliary verb in a VPAux consists of a subject agreement morpheme, a root and an ending in basic form. The customary ending **-a** can in certain cases be **-ê**, **-e**, or **-ô**. Auxiliary verbs and main verbs, therefore have the same basic morphological structure. Auxiliary verbs do not take object agreement morphemes but some can take the progressive morpheme **sa**, the potential morpheme **ka**, the future tense morpheme **tla**, or the present tense morpheme **a**. In (4) the progressive morpheme **sa** is prefixed to **-ntse**.

(4) Ke sa ntse ke ja.

ke-sa-ntse	*ke-j-a*
AgrSubjP1sg-ProgPre-Aux	AgrSubjP1sg-eat-VEnd

I am still eating.

3.2.2 Syntactic characteristics

The minimal structure of a VPAux consists of the auxiliary verb and an obligatory complement, viz. the VPMain, VPidcop, VPDescop, VPAsscop or another VPAux (Berg, 2018 119). In (5) the auxiliary verb **o ne** is followed by the VPMain **a reka**. In (6) the auxiliary verb **re setse** is followed by the VPMain **re kwala ditlhatlhobo**. In (7) the auxiliary verb **re ne** is followed by the VPAux **re tlhola re ya kwa lewatleng ka selemo**.

(5) O ne a reka.

o-ne	*a-rek-a*
AgrSubj1-Aux	AgrSubj1-buy-VEnd

She was buying.

[2] The category of improper auxiliaries in Diagram 1 is based on the syntactic features of these auxiliaries. They can act as verbs as well as auxiliaries. They do not act as modal verbs when used as auxiliaries in Setswana. When compared to the modal verbs identified by Mchombo (2004:30-32) for Chichewa the only auxiliary verb that may be used as modal for Setswana is the proper auxiliary **-bô** with coordinative consecutive value with the idea of continuing something.

(6) Re setse re kwala ditlhatlhobo.
re-setse *re-kwal-a*
AgrSubjP1pl-Aux AgrSubjP1pl-write-VEnd
di-tlhatlhob-o
NPre10-examine-DevSuf
We are already writing examinations.

(7) Re ne re tlhola re ya kwa lewatleng ka selemo.
re-ne *re-tlhola* *re-y-a*
AgrSubjP1pl-Aux AgrSubjP1pl-Aux AgrSubjP1pl-go-VEnd
kwa *le-watle-ng* *ka* *se-lem-o*
LocPart NPre5-sea-LocSuf TempPart NPre7-plough-DevSuf
We always went to the seaside in summer.

The main predicate in a Setswana sentence is always a main or copulative verb even if these verbs are the complements of auxiliary verbs. Auxiliary verbs contribute polarity (refer to (12)), tense, aspect and time[3] features (cf. Diagram 1) to the f-structure of the main verb. In terms of subcategorisation frames, there are two main approaches to analyse auxiliaries in LFG. The auxiliary verb is treated as a special type of raising verb that takes a SUBJ and an XCOMP[4] argument, or is considered a feature-carrying element (Butt *et al.*, 1999:61-63; Falk, 2003). Butt *et al.* (1996) treat auxiliaries as feature carrying elements, rather than as raising verbs. They state that "this avoids unnecessary structural complexity and provides a uniform cross linguistic analysis which eases the burden for machine translation" (Butt *et al.*, 1996:111). We treat Setswana auxiliary verbs as feature-carrying elements. The auxiliary verb does not have a subcategorisation frame and a flat f-structure analysis is followed because the complement following the auxiliary verb is considered the main predicate of the sentence (Berg, 2018:138).

In (8) the Setswana sentence includes an auxiliary verb **ba ne**. The f-structure of this auxiliary verb is presented in Figure 1. The f-structure of the complement **ba reka ditlhako** (they buy shoes) is presented in Figure 2. The f-structure of (8) is presented in Figure 3. The auxiliary verb indicates the relative past tense while the subject agreement morpheme **ba** indicates class 2 agreement.

[3] Tense contributes to the substantiating of time, and therefore the verbal markedness of time and tense is often identical. The major device in the establishing of time is the interrelation between tense forms and time adverbials, which indicates that time interpretation is primarily constituted by factors outside the verb. The role of the tense-related auxiliary verbs lies in the indication of the relation between coding time and reference time. The auxiliary verb is thus only remotely related to the indication of time (Pretorius, 1997:166)
[4] The XCOMP function is as an open clausal function that does not contain an internal subject phrase (Dalrymple, 2001:24).

(8) Ba ne ba reka ditlhako.

ba-ne	*ba-rek-a*	*di-tlhako*
AgrSubj2-Aux	AgrSubj2-buy-VEnd	NPre8-shoe

They were buying shoes.

```
"ba ne"
   [TNS-ASP [TENSE relpast]]
  1[AUX +, CLASS 2        ]
```

Figure 1: f-structure of ba ne

```
"ba reka ditlhako"
    [PRED    'REK<[1-SUBJ:PRO], [15:tlhako]>']
    [SUBJ    [PRED 'PRO']                   ]
    [           [PRED 'tlhako'      ]       ]
    [OBJ     15 [CLASS 8, NTYPE ord ]       ]
    [TNS-ASP [MOOD indicative]              ]
   1[CLASS 2, VTYPE main                    ]
```

Figure 2: f-structure of ba reka ditlhako

```
"ba ne ba reka ditlhako"
    [PRED    'REK<[1-SUBJ:PRO], [29:tlhako]>']
    [SUBJ    [PRED 'PRO']                   ]
    [           [PRED 'tlhako'      ]       ]
    [OBJ     29 [CLASS 8, NTYPE ord ]       ]
    [TNS-ASP [MOOD indicative, TENSE relpast]]
   1[AUX +, CLASS 2, VTYPE main             ]
```

Figure 3: f-structure of (8)

When multiple auxiliary verbs appear in the same phrase they have the following order:

negative (structures) > tense > aspect > time

3.2.3 Semantic characteristics

Auxiliary verbs collaborate with and contribute to the meaning of the main verb and the range of categories they express in their morphology. Cole (1955:235) states that auxiliary or deficient verbs “are used both to indicate simple distinctions of time or tense and to provide special implications or modifications of significance as distinct from or additional to the time factor”. In this regard Louwrens (1994:17) states that auxiliary verbs enrich the meaning of their complementary verbs adding semantic information regarding the progression or completion of an action. Based on the semantic groupings in Diagram 1 we add duration as aspect in addition to progressive, persistive and perfective which are already part of f-structure. However, the semantic values of many auxiliary verbs indicate a temporal value which has not been linked to an aspectual category but should be expressed at f-structure. We propose a solution in 4.3.4. As we focus on the tense auxiliaries we now briefly contextualise a framework for discussion.

4. Tense

4.1 Tense analysis

Reichenbachs' (1947) well known framework for the analysis of tense/aspect (T/A) employs the relationship between event time (E), speech time (S) and reference time (R) to determine and categorise T/A. The relation between R and S is taken to code tense while the relation between E and R encodes aspect (Butt & Rivzi, 2008:60). Reichenbachs' original work has been revised in several versions since reaching agreement that the relationship between E and S is indirect and mediated by R (Kamp & Reyle, 1993; Hornstein, 1990; Butt & Rivzi, 2008).

4.2 Tense and aspect in Bantu languages

Nurses' (2008) contribution to the discussion of T/A in the Bantu languages is comprehensive and includes data from more than 200 Bantu languages and varieties. He discusses tense systems evidenced in non-relative affirmatives. In the statement of his conceptual framework Nurse (2008:10-15) gives the following guidelines regarding T/A in Bantu languages:

- Tense and aspect is a system;
- Tense and aspect systems are cognitively based, not direct representations of events in the real world;
- Tense and aspect form an interlocking system;
- A discreet verbal tense aspect form has a specific and unique range of meaning;
- The system is not inflexible or unchanging;
- Every finite verb form has aspect; and
- Most Bantu languages encode tense on the left and aspect on the right.

These guidelines are all applicable to tense and aspect in Setswana.

4.3 Tense in Setswana

The treatment of tense in standard Setswana grammars is limited and has been presented from different perspectives. Earlier descriptions of tense in the Sotho languages (Sharpe, 1952; Sandilands, 1953; Cole, 1955; Doke & Mofokeng, 1957; Lombard *et al*., 1985) distinguished between simple and compound tenses based on the components of the verb phrase. The simple tenses were also referred to as mono-verbal tenses. Compound tenses were identified when auxiliary verb phrases appear. The classification is based on purely formal grounds rather than on the semantic values and usage of these tenses. The problem with this classification was that two tenses cannot be identified in a single auxiliary verb phrase which was the result of not noticing the aspectual value contributed by the auxiliary verb.

Later studies follow Reichenbach's interpretation of tense as a deictic category (Pretorius 1997, 2003, Krüger, 2013a, 2013b). Setswana, as do most other languages, uses two strategies to locate events in time with reference to speech time (S) (the deictic centre). The first is to vary the morphological form of the verb which is *tense marking*, and the second is to use *temporal adverbs*. Two types of tenses are generally distinguished namely absolute and relative tenses distinguishing a past, present and future tense. In the absolute tense forms "the reference point from which the temporal semantic interpretations are done, is the deictic centre (coding time)" (Posthumus, 1990:23). The grammatical tense forms of anteriority, simultaneity and posteriority, are generally referred to as past (events prior to coding time) (9), present (events coinciding with coding time) (10) and future tense (events subsequent to coding time) ((11) and (12)) respectively.

Example (9) is in the past tense: S is now, R is now, and E is some time in the past. Note that the perfective suffix here contributes past tense in the absolute tenses while it contributes perfective aspect in inchoative verbs in relative tenses. Example (10) is in the present tense: S is now, R is now and E is now. Examples (11) and (12) are in the future tense: S is now, R is now, and E is some time in the future.

(9) Mosadi o rekile ditlhako.
mo-sadi *o-rek-il-e* *di-tlhako*
NPre1woman AgrSubj1-buy-PerfSuf-VEnd NPre8-shoe
The woman bought shoes.

(10) Mosadi o reka ditlhako.
mo-sadi *o-rek-a* *di-tlhako.*
NPre1woman AgrSubj1-buy-VEnd NPre8-shoe
The woman buys / is buying shoes.

(11) Mosadi o tla reka ditlhako.
mo-sadi *o-tla-rek-a* *di-tlhako*
NPre1-woman AgrSubj1-FutPre-buy-VEnd NPre8-shoe
The woman will buy shoes.

We include the negative of (11) in (12) to show the negative polarity contributed by the negative morpheme **ga** in the auxiliary verb **a ketla** (Figure 4 and Figure 5).

(12) Mosadi ga a ketla a reka ditlhako.
mo-sadi *ga-a-ketla* *a-rek-a*
NPre1-woman NegPre-AgrSubj1-Aux AgrSubj1-buy-VEnd
di-tlhako
NPre8-shoe
The woman will not buy shoes.

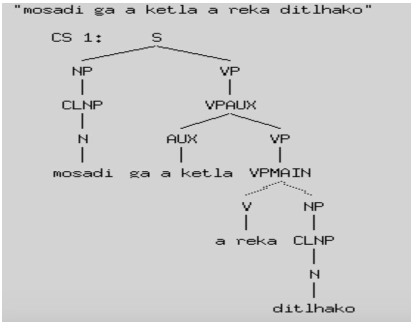


Figure 4: c-structure of (12)

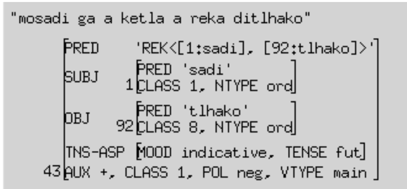


Figure 5: f-structure of (12)

In the relative tense forms the interpretation of E is made in relation to R, which is in turn established in relation to S which is constituted by the deictic centre. In Setswana, the form of the auxiliary verb in the auxiliary verb phrase expresses the relation between S and a newly established R.

4.3.3 Paradigm of Setswana tenses

The following tense paradigm (Table 1) for Setswana was developed based on Nurse (2007b; 2008):

	Perfective	**Imperfective**	**Progressive**	**Persistive**
Past	Re rekile. (We bought.) Re ne re rekile. (We had bought.)	Re ne re reka. (We were buying.) Re ne ra reka. (We bought. (subsequently))	Re ne re sa ntse re reka. (We were still buying.)	Re ne re ntse re reetsa. (We had been listening.)
Present		Re reka. (We buy.) Re a reka. (We are buying.)	Re sa ntse re reka. (We are still buying.)	Re ntse re reetsa. (We have been listening.)
Future	Re tlabo re rekile. (We will have bought.)	Re tla reka. (We will buy.) Re tlabo re reka. (We will be buying.)	Re tlabo re sa ntse re reka. (We will still be buying).	Re tlabo re ntse re reetsa. (We will be listening.)

Table 1: Setswana tense paradigm

Setswana tenses are marked in the morphology of the verb. The auxiliary verbs **o ne** (13) and **o tlabo** (14), indicate the shifting of R to a point prior to or subsequent to S respectively. Their complement **a reka** contributes an imperfective value even though it is not overtly marked.

(13) Mosadi o ne a reka ditlhako.

mo-sadi	*o-ne*	*a-rek-a*	*di-tlhako*
NPre1-woman	AgrSubj1-Aux	AgrSubj1-buy-VEnd	NPre8-shoe

The woman was buying shoes.

(14) Mosadi o tlabo a reka ditlhako.

mo-sadi	*o-tlabo*	*a-rek-a*	*di-tlhako*
NPre1-woman	AgrSubj1-Aux	AgrSubj1-buy-VEnd	NPre8-shoe

The woman will be buying shoes.

4.3.4 Features of auxiliary verbs following tense-indicating auxiliary verbs

The auxiliary that follows **-ne** or **-tlabo** is aspectual or it may contribute a temporal value. In (15), (16) and (17) the auxiliary verb **ke tlabo** indicates the relative future tense. In (15) the auxiliary verb **ke santse** indicates persistive aspect while the main verb **ke bala**, even though it is not overtly marked, indicates imperfective aspect. In (16) the auxiliary verb **ke setse** expresses logical time with the adverbial value of *already*. In (17) the main verb **ke badile** indicates perfective aspect.

(15) Ke tlabo ke santse ke bala.

ke-tlabo	*ke-santse*	*ke-bal-a*
AgrSubjP1sg-Aux	AgrSubjP1sg-Aux	AgrSubjP1sg-read-VEnd

I will still be reading.

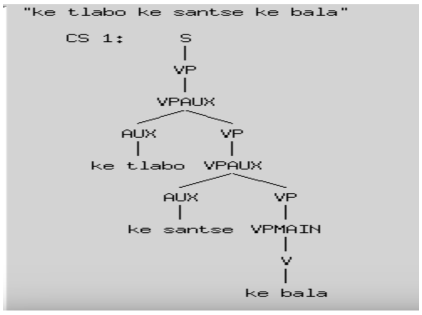


Figure 6: c-structure of (15)

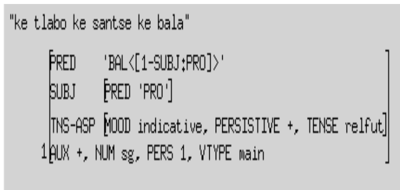


Figure 7: f-structure of (15)

(16) Ke tlabo ke setse ke bala.

ke-tlabo	*ke-setse*	*ke-bal-a*
AgrSubjP1sg-Aux	AgrSubjP1sg-Aux	AgrSubjP1sg-read-VEnd

I will already be reading.

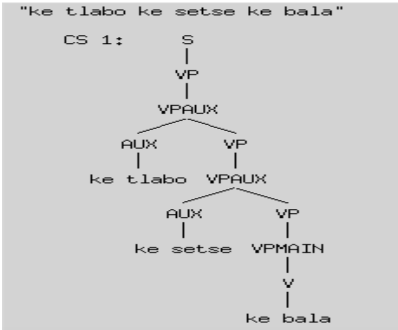


Figure 8: c-structure of (16)

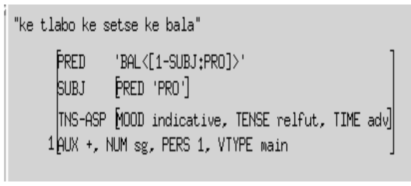


Figure 9: f-structure of (16)

(17) Ke tlabo ke setse ke badile.
ke-tlabo *ke-setse*
AgrSubjP1sg-Aux AgrSubjP1sg-Aux
ke-bal-il-e
AgrSubjP1sg-read-PerfSuf-VEnd
I will already have read.

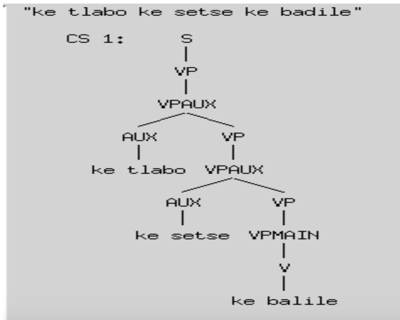


Figure 10: c-structure of (17)

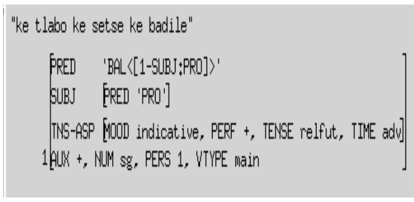


Figure 11: c-structure of (17)

In (18), (19), (20) and (21) the auxiliary verb **ba ne** indicates relative past tense. In (18) the auxiliary verb **ba santse** indicates persistive aspect. In (19) the auxiliary verb **ba setse** expresses logical time with the adverbial value of already. In (20) the main verb **ba re thusitse** indicates perfective aspect. In (21) three auxiliary verbs are employed, where **ba tlhola** expresses a durative aspect.

(18) Banna ba ne ba santse ba re thusa.
ba-nna *ba-ne* *ba-santse*
NPre2-man AgrSubj2-Aux AgrSubj2-Aux
ba-re-thus-a
AgrSubj2-AgrObjP1pl-help-VEnd
The men were still helping us.

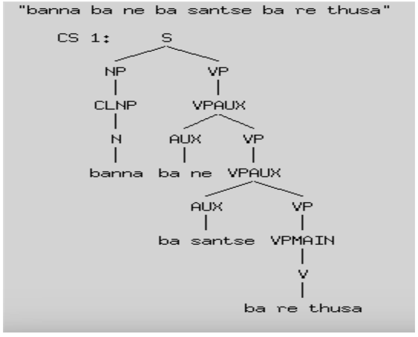


Figure 12: c-structure of (18)

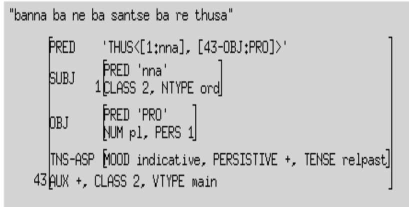


Figure 13: f-structure of (18)

(19) Banna ba ne ba setse ba re thusa.
ba-nna *ba-ne* *ba-setse*
NPre2-man AgrSubj2-Aux AgrSubj2-Aux
ba-re-thus-a
AgrSubj2-AgrObjP1pl-help-VEnd
The men were already helping us.

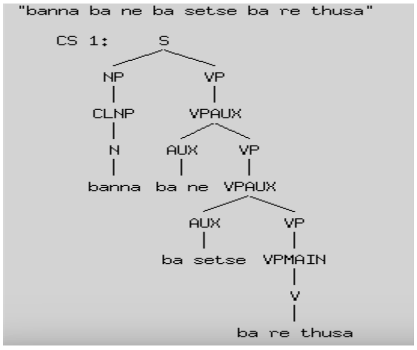


Figure 14: c-structure of (19)

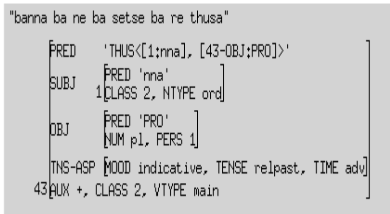


Figure 15: f-structure of (19)

(20) Banna ba ne ba setse ba re thusitse.
ba-nna *ba-ne* *ba-setse*
NPre2-man AgrSubj2-Aux AgrSubj2-Aux
ba-re-thus-il-e
AgrSubj2-AgrObjP1pl-help-PerfSuf-VEnd
The men had already helped us.

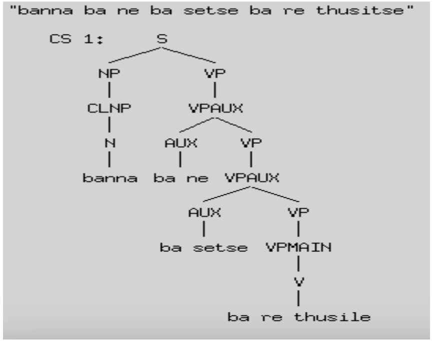


Figure 16: c-structure of (20)

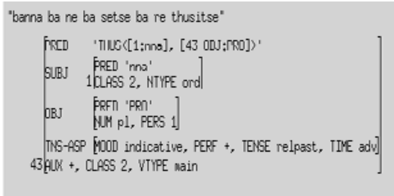


Figure 17: f-structure of (20)

(21) Banna ba ne ba santse ba tlhola ba re thusa.
ba-nna *ba-ne* *ba-tlhola* *ba-santse*
NPre2-man AgrSubj2-Aux AgrSubj2-Aux AgrSubj2-Aux
ba-re-thus-a
AgrSubj2-AgrObjP1pl-help-VEnd
The men were still always helping us.

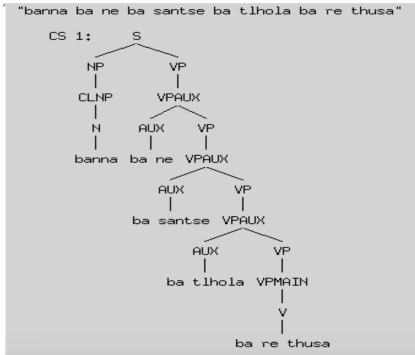


Figure 18: c-structure of (21)

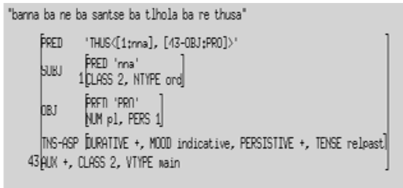


Figure 19: f-structure of (21)

Setswana auxiliaries presented as examples in this paper always add the same features regardless of other auxiliaries in the phrase, they are therefore purely agglutinative. However, as is shown in example (22) there is doubt whether this is true in all cases and more detailed work is needed to clarify the matter.

When the tense indicating auxiliary **-ne** is followed by tense indicating **-tlabo** as in (22), **-ne** indicates tense while **-tlabo** now indicates a probability and not tense anymore. A single verb cannot have multiple tenses.

(22) Ke ne ke tlabo ke santse ke dira.

ke-ne	*ke-tlabo*	*ke-santse*
AgrSubjP1sg-Aux	AgrSubjP1sg-Aux	AgrSubjP1sg-Aux

ke-dir-a
AgrSubjP1sg-work-VEnd
I would still have been working.

It is interesting to note that **-tlabo** seems to be used interchangeably with **-kabo**. In these auxiliary verbs the potential morpheme **ka** and the future tense morpheme **tla** combine with the inferential/stative auxiliary **-bo**, indicating semantic correspondence between potential and future events.

6. Summary

In expanding the Setswana LFG grammar we have added auxiliary verbs with their semantic values to the lexicon. More work is still needed though. Auxiliary verbs appear juxtaposed but then in a set order when more than one auxiliary verb appear in the same VP. Auxiliary verbs indicating tense may be followed in a semantically compositional manner by auxiliary verbs indicating aspect and time.

We extend aspectual values in the grammar to include further information about the nature of the event and add *durative* aspect for the newly added auxiliary verb **-tlhola** which has the value of duration/continuation in Diagram 1 (cf. Butt & Rizvi, 2008:60). Several other auxiliary verbs in Diagram 1, such as **-setse** have a value that is not one of the known aspects. They express the semantic value of "logical time" which we propose to link to the attribute *time* with the value of adverb (adv) at f-structure. When translated into English **-setse** has the meaning of *already*. Time is a generic attribute in Setswana f-structure and other auxiliary verbs with this adverbial value could in the interim be treated in this manner. Further work is still needed to consider telic and itive aspect (cf. Nurse, 2007a:164-165).

References

Berg, Anna S. 2018. Computational syntactic analysis of Setswana. Potchefstroom: North-West University. (Thesis - PhD).

Brits, Karien, Pretorius, Rigardt and Van Huyssteen, Gerhard B. 2005. Automatic lemmatization in Setswana: towards a prototype. *South African Journal of African Languages*, 25(11):37-47.

Butt, Miriam, Niño, María-Eugenia and Segond Frédérique. 1996. Multilingual Processing of Auxiliaries within LFG. In: Gibbon David. ed. *KONVENS '96. Natural Language Processing and Specch Technology: Results of the 3rd KONVENS Conference*. Mouton de Gruyter, 111-122. https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/.pdf

Butt, Miriam, King, Tracey H, Niño, María-Eugenia and Segond Frédérique. 1999. *A grammar writer's cookbook*. Stanford: CSLI.

Butt, Miriam and Rizvi, Jafar. 2008. Tense and aspect in Urdu. In: Hofherr, Patricia C. and Laca, Brenda. eds. *Layers of aspect*. California: CSLI publications. p.43-62.

Cole, Desmond T. 1955. *An introduction to Tswana grammar*. Cape Town: Longman.

Doke, Clement M. and Mofokeng, Sophonio M. 1957. *Textbook of Southern Sotho grammar*. Cape Town: Longmans & Green.

Falk, Yehuda. 2003. The English auxiliary system revisited. In Butt Miriam and King Tracey H. eds. *Proceedings of the LFG03 Conference*. CA: CSLI Publications, 184-204. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/8/pdfs/lfg03.pdf

Hornstein, Norbert. 1990. *As time goes by*. Cambridge: MIT Press.

Kamp, Hans. and Reyle, Uwe. 1993. *From discourse to logic*. Dordrecht: Kluwer Academic Publishers.

Krüger, Caspar J.H. 2006. *Introduction to the morphology of Setswana*. München: Lincom Europe.

Krüger, Caspar J.H. 2013a. *Setswana syntax: a survey of word group structures: Volume 1*. München: Lincom Europe.

Krüger, Caspar J.H. 2013b. *Setswana syntax: a survey of word group structures: Volume 2*. München: Lincom Europe.

Lombard, Daan J., Van Wyk, Egidius B. and Mokgokong, Pothinus C. 1985. *Introduction to the grammar of Northern Sotho*. Pretoria: Van Schaik.

Louwrens. Louis J. 1994. Aspect as an essential sub-category of the verb in Northern Sotho. *Journal for African Languages*, 14(3):116-127.

Maho, Jouni F. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In: Nurse, Derek and Philippson, Gerard. eds. *The Bantu Languages*. London: Routledge, 639-651.

Mchombo, Sam A. 2004. *The syntax of Chichewa*. Cambridge: Cambridge University press.

Nurse, Derek. 2007a. The emergence of tense in early Bantu. In: Payne, Doris L. and Peña, Jaime. eds. *Selected Proceedings of the 37-th Annual Conference on African Linguistics*. Somerville: MA, 164-179. http://www.lingref.com/cpp/ acal/37/paper1604.pdf

Nurse, Derek. 2007b. *Appendix 1: Matrices and Notes for Matrix Languages*. http://www.ucs.mun.ca/~dnurse/tabantu.html.

Nurse, Derek. 2008. *Tense and aspect in Bantu*. Oxford: Oxford University Press.

Posthumus, Lionel C. 1990. Time reference in Zulu. *South African Journal for African Languages*, 10(1):22-28.

Pretorius, Laurette, Viljoen, Biffie, Pretorius, Rigardt and Berg, Ansu. 2009. A finite-state approach to Setswana verb morphology. In: Yli-Jyrä Anssi, Kornai András, Sakarovitch Jacques and Watson Bruce. Eds. *Finite-State Methods and Natural Language Processing* Berlin: Springer, 131-138).

Pretorius, Rigardt S. 1997. Auxiliary verbs as a subcategory of the verb in Tswana. Potchefstroom: Potchefstroom University for Christian Higher Education. (Thesis - PhD).

Pretorius, Rigardt S. 2003. Absolute tense forms in Tswana. *Journal for language teaching*, 37(1):13-25.

Pretorius, Willem J. 2004. ‘n Dimensionele grammatiese beskrywing van hulpwerkwoorde in Noord-Sotho. Johannesburg: Randse Afrikaanse Universiteit. (Thesis - D.Litt. et Phil).

Ranamane, Tlhabane D. 2009. Discourse functions of tense and aspect in Setswana narrative texts. Pretoria: University of South Africa. (Thesis - PhD).

Reichenbach, Hans. 1947. *Elements of symbolic logic*. New York: The Free Press.

Sandilands, Alexander. 1953. *Introduction to Tswana*. Morija: Morija Printing Works.

Setshedi, Jacob E. 1974. The auxiliary verbs and the deficient verbs in Tswana. Pietersburg: University of Natal. (Dissertation - MA).

Sharpe, Maud R.L. 1952. *Everyday Sesotho Grammar*. Morija: Morija Sesuto Book Depot.

Assamese Case Alignment Shifts in Progress

Pori Saikia
University of Essex

Maris Camilleri
University of Essex

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 251–271

http://csli-publications.stanford.edu/LFG/2019

Keywords: Indo-Aryan languages, split ergative, differential case marking, language change

Saikia, Pori, & Camilleri, Maris. 2019. Assamese Case Alignment Shifts in Progress. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 251–271. Stanford, CA: CSLI Publications.

Abstract

This paper looks at case alignment in Assamese from both a synchronic and diachronic point of view. We take the task of tracing the development of the ergative case marker from the language's proto-period, and see how it evolved. This study, for the first time, provides a comparison of adult and child language data. Beyond the account of Assamese as a split ergative language, our study's results show that the semantic factor guiding this split is changing. From an ergative system based on split intransitivity determined by agentivity, thus realizing a split between unergative and unaccusative SUBJs, ERG marking is emerging on the ANIMATE subjects of unaccusatives. Interestingly, we also find that there are already traces of evidence of Assamese having had possibly the onset of an ANIMACY-based subject marking distinction in its proto-period.

1 Introduction

In this paper we determine that Assamese, an Indo-Aryan language spoken by 14 million native speakers in the northeastern state of Assam in India, is showing signs of change in its morphosyntax. Specifically we argue that the language is reanalysing its differential subject marking system guided by semantically-motivated case alternations that are changing the nature of the current status of the language which is one of split ergativity based on agentivity. In effect, we argue that what is taking place in Assamese very much parallels the situation in other New Indo-Aryan (NIA) languages, as argued in, for e.g. Ahmed (2010) and Butt & Ahmed (2011), where the language is recycling its current SUBJ case system to express distinct semantic factors.

Supporting the thrust of this study which is a discussion of a hypothesis that change is in progress, we incorporate a child language data-based study from Saikia (in prep.) and demonstrate that the direction in which change is progressing is magnified by what can be observed through child language data. The employment of child language data as a means which can guide our assessments on, and of, variation and change, is key to the views upheld in Lightfoot (2010). We also argue that the newly evolving split is conditioned by the semantic nature of the NPs, and which is in fact a reflection of the differential marking one finds with respect to objects in the language. Consequently we hypothesise that what is emerging can also be referred to, in parallel, as differential subject marking.

[†]We thank ESRC for partially funding this project. We are also grateful to Mr. Dennis Somadula for making the illustrations of the elicitation task pictures, and Mr. Rocktim Gohain for assisting us with the fieldwork. We also thank the participants of LFG 2019 and the anonymous reviewers for their feedback and input.

[†]Abbreviations used: ACC: accusative CLF: classifier; DAT: dative; ERG: ergative; F: feminine; GEN: genitive; IMP: imperative; INS: instrumental; M: masculine; NOM: nominative; PRES: present; PROG: progressive; PERF: perfect; PST: past; PASS: passive; PTCP: participle; PL: plural; SG: singular

This paper is organised as follows. In section 2 we provide a characterisation of case marking in Assamese. In section 3 we integrate the child language data and its corresponding adult data studies to our overall assessment of the language's grammar and pinpoint the change in progress. In section 4 we then provide a summary of our conclusions.

2 Distribution of case in Assamese

2.1 Assamese as a split ergative language

Assamese is a head-final SOV dominant language that is syntactically accusative, i.e. its different subjects, as we will exemplify later, align in one pattern together with respect to control phenomenon, anaphoras, relativisation, and in particular agreement patterns. When it comes to the morphological characterisation of the case system, several terminologies to code case alignments are provided in the literature (cf. (Dixon, 1979, 1994; Comrie, 1978)), however, we specifically choose the terminologies used in Mohanan (1994), and refer to Assamese as a split ergative language, i.e. a language with two distinct cases associated with subjects, where one is inflected and the other remains uninflected or unmarked. The marked subject is referred to as being ergative, while the latter nominative.

Assamese is often mischaracterised as a NOMINATIVE-ACCUSATIVE (Kakati, 1941; Goswami & Tamuli, 2003; Nath, 2003; Haddad, 2011) or a (fully) ERGATIVE system (Devi, 1986; Butt & Deo, 2001; Zakharyin, 2015). However, what Assamese really demonstrates is a **split ergative system** with splits conditioned by **intransitivity**, i.e. based on whether the intransitive verb is unergative or unaccusative, which, synchronically, without yet considering the direction of the change in progress, is based purely on **agentivity** (Amritavalli & Sarma, 2002). Within the pronominal system, however differential case marking (DCM) (Aissen, 1999, 2003) is conditioned by PERSON and NUMBER (Saha & Patgiri, 2013).

To understand why we are referring to Assamese as a split ergative language, we provide the data below. NP subjects of (di)transitive verbs (A), irrespective of animacy, obligatorily take an overt ERG case marker in Assamese, as exemplified through the data in (1).[1]

(1) a. lora-tu=**e** bol-tu d^hor-i as-e
boy-CLF=ERG ball-CLF hold-PROG be.PRES-3
'The boy is holding the ball.'
(ANIM ERG SUBJ **of transitive** PRED)

b. bas-bur=**e** baik-k^hɔn k^hundi-a-l-e
bus-PL=ERG bike-CLF knock down-CAUS-PST-3
'The buses knocked down the motorbike.'
(INANIM ERG SUBJ **of transitive** PRED)

[1]The Assamese data, unless provided with a citation, is the native speaker author's own.

Among intransitive verbs, agent-like subjects of unergatives (S_a), irrespective of animacy, trigger an overt marker on the subject, while the patient-like subjects of unaccusative verbs (S_o) remain unmarked. For example, the S_a NP referent of unergative verbs like *jump*, *dance*, and *swim* control an activity, as opposed to the S_o NP referents of unaccusative verbs like *fall*, *sink*, and *burn* that have no control over the activity. Further, similar to referents of an O function, the referents of S_o could be affected by the event. Although certain intransitive verbs can be easily categorised as either S_a or S_o, the categorisation of some might vary across languages (see, for instance, Dixon (1979); van Valin Jr (1990); Handschuh (2008)). The contrast between S_a and S_o, in the context of unergative and unaccusative verbs respectively, is illustrated through (2) and (3).

(2) a. roza-zon=**e** xãtur-i as-e / xãtur-is-e
king-CLF=ERG swim-PROG be.PRES-3 / swim-PERF-3
'The king is / has been swimming.'
(ANIM ERG **NP** SUBJ **of unergative** PRED)

b. botah-zak=**e** huhurija-is-e
wind-CLF=ERG whistle-PERF-3
'The wind has been whistling.'
(INANIM ERG **NP** SUBJ **of unergative** PRED)

(3) a. roza-zon.∅ boh-i as-e / boh-is-e
king-CLF.NOM sit-PROG be.PRES-3 / sit-PERF-3
'The king is / has been sitting.'
(ANIM NOM **NP** SUBJ **of unaccusative** PRED)

b. kat^h-sota.∅ upoŋ-i as-e / upoŋ-is-e
wood-CLF.NOM float-PROG be.PRES-3 / float-PERF-3
'The piece of wood is / has been floating.'
(INANIM NOM **NP** SUBJ **of unaccusative** PRED)

Case marking is more complex in the pronominal system. On the basis of the discussion by Saha & Patgiri (2013), specifically in Assamese, only the 2nd and 3rd PERSON **plural pronouns** trigger ERG case marking in the form of an enclitic.[2]

(4) a. tumaluk=**e** xãtur-i as-a / xãtur-is-a
2.PL=ERG swim-PROG be.PRES-2 / swim-PERF-2
'You (PL) are / have been swimming.'
(**2**.PL ERG **pronoun** SUBJ **of unergative** PRED)

[2] In another Indo-Aryan language, Punjabi, the pronominal system appears to be sensitive *just* to 1st/2nd vs. 3rd PERSON based split, whereby only the latter set of pronouns (and NPs) take an ERG marking (Butt & Deo, 2001). This, thus differs from the seemingly more complex interaction between PERSON and NUMBER in Assamese.

b. xĩhot=**e** xãtur-i as-e / xãtur-is-e
3.PL=ERG swim-PROG be.PRES-3 / swim-PERF-3
'They are / have been swimming.'
(**3.**PL ERG **pronoun** SUBJ **of unergative** PRED)

Supposedly, the rest of the pronominal SUBJ paradigm remains unmarked, i.e. it expresses NOM case, (as the ∅ marking is meant to illustrate) demonstrated via the 1.SG and 3.SG.M pronominal subject forms in (5). In this respect, therefore, the split on the basis of PERSON and NUMBER within the pronominal system, in contrast to the neater nominal system, takes supremacy over the requirement of (A) SUBJs of transitive predicates to be ERG-marked as illustrated in (1).[3]

(5) a. moi.∅ sur-tu=**k** d^hor-il-u
1.SG.NOM thief-CLF=ACC hold-PST-1
'I caught the thief.'
(**1.**SG NOM **pronominal** SUBJ **of transitive** PRED)

b. xi.∅ sur-tu=**k** d^hor-il-e
3.SG.M.NOM thief-CLF=ACC hold-PST-3
'He caught the thief.'
(**3.**SG.M NOM SUBJ **of transitive** PRED)

An internal reviewer suggests that there is a probability that the observation of unmarked pronominal forms may look so only on the basis of their surface morphology, i.e. in the absence of an *-e* marking. For this reason, an alternative analysis would be to assume that these pronouns are in fact 'old and have come down (for some reason) in an originally oblique form'. In support of this alternative analysis, we could argue, following Kakati (1941), that the PERSON and NUMBER based split in the pronominal system is itself a remnant from Middle Indo-Aryan (MIA). He observes how for instance, the 1.SG pronoun *moi* or the inferior 2.SG *toi*, and so on maintain the MIA proto-instrumental forms *-ẽ*, *-i* (synchronically interpreted as ERG) in their extended oblique pronominal bases. On the other hand, (Saha & Patgiri, 2013, pp. 39-40) argue that the split that results is a reflex of a morphophonological constraint, such that since the 1.SG/PL, 2.SG and 3.SG pronominal forms end with a high vowel */i/*, ERG *-e* marking is blocked.

Given the above characterisation for the SUBJ case marking system for nominals, inclusive of a split intransitivity governed by the subject's agentivity, along with an incorporation of the assumption that the pronominal system is actually characterised by DCM based on PERSON and NUMBER, the following table summarises the facts.[4]

[3]The same parallel behaviour follows for S_a subjects.

[4]It is worth mentioning that pronouns in Assamese have always been discussed with respect to animate reference; the distribution of which, in terms of case marking, is presented in Table (1). Reference to inanimate entities, on the other hand, involves a distinct pronominal device; a resort to the use of the demonstrative pronominal paradigm, such as *ei*/*hei* 'this/that' along with the attachment of the default classifier *-tu*, or any of the shape classifiers, such as *-dal* and *-k^hɔn*.

SUBJ	NP	1.SG/PL	2/3.SG	2/3.PL
A	ERG	NOM	NOM	ERG
S_a	ERG	NOM	NOM	ERG
S_o	NOM	NOM	NOM	NOM

Table 1: Distribution of case-marking on SUBJ GFS

Another instance where the ergative split discussed above for NPs as well as pronouns is overridden, is in contexts where a homophonous *-e* marker is present on SUBJs to express what Butt & Holloway King (1991), Butt (2006), Ahmed (2010), and Butt & Ahmed (2011) refer to, with respect to cognate *-ne* marker in Hindi/Urdu, as a marker of volitionality/intentionality. If we consider the contrast in the pair below, *lora* 'boy' is unmarked in the context of the intransitive unaccusative verb *por* 'fall', in line with our discussion above. Nevertheless, when the semantic interpretation expressed is such that the SUBJ deliberately/purposefully initiates the falling event, then an *-e* marker, which we presume to be the ERG marker, but which here is of a distinct semantic function is marked on the SUBJ. The end result is such that as illustrated in (6b), we have the presence of an ERG marker in the context of an unaccusative verb.

(6) a. lora-tu.∅ por-il(-e)
boy-CLF.NOM fall-PST(-3)
'The boy fell down.'

b. lora-tu=**e** por-i di-l-e
boy-CLF=ERG fall-NF give-PST-3
'*The boy* (deliberately/purposefully) fell down.' (Chowdhary, 2014, p. 111)

Saha & Patgiri (2013, p. 40) argue that the same follows for pronouns, including the 1st PERSON.SG/PL and 2nd/3rd.SG pronouns, where NOM marking is overridden, and *-e* marking is present, as illustrated through the 1.SG pronominal *moi* 'I' in (7).[5]

(7) moi=**e** za-m tumar log-ot
1.SG=ERG go-FUT 2.SG.GEN company-LOC
'It is *I*, who will accompany you' (S.C. Chiring Phukan, p.c.)

Before proceeding further, a note on agreement behaviour vis-à-vis case is in order. In Assamese, NOM vs. ERG DCM on SUBJs does not block agreement with the subject on verbs, unlike what goes on in other Indo-Aryan ergative languages,

[5]Note that this particular use of the ERG form on the 1.SG pronominal can also be realised as *-ei* or *-(e)he*. The availability of alternations of this sort is also true for the 1.PL and 2/3.SG.

such as Hindi. This is exemplified, for instance, through data such as (4a) and (5a), where irrespective of ERG vs. NOM marking, respectively, the verb displays the relevant PERSON agreement with it. Notwithstanding this pattern, DCM *does* matter for SUBJ-verb agreement purposes beyond ERG/NOM-marked SUBJ contexts. Non-canonical subjects, which are expressed via non-ERG/NOM morphology, on the other hand trigger default 3rd PERSON agreement. In (8), we have an illustration of *lag* lit. 'want' functioning as a psych predicate, with the meanings 'feel' and 'get (fear)', respectively. This consequently requires a GEN-marked SUBJ, such as the 1.SG pronominal *mur*. The presence of such a subject triggers 3rd PERSON agreement on the verb, which is the form employed for default agreement contexts. A similar default agreement pattern also follows in the case of the predicate *lag* when used in a desiderative sense, meaning 'want', as in (9). In this case, the SUBJ is ACC/DAT-marked via the phonologically-conditioned allomorphs *-(o)k*.[6]

(8) a. mur ijat niz=ok asohua zen lag-e
1.SG.GEN here self=ACC stranger as if get.PRES-3
'I feel as if I am a stranger here.' (Chowdhary, 2014, p. 115)

b. mur b^{h}oi lag-is-e
I.SG.GEN fear get-PERF-3
'I am scared' (Lit: 'I have got fear.')
(GEN SUBJ of psych verb)

(9) muk b^{h}at lag-e
1.SG.DAT/ACC rice want.PRES-3
'I want rice.'
(DAT/ACC SUBJ of desiderative 'want')

2.2 The historical development of the Ergative in Assamese

Ergativity was not an inherent grammatical feature of Sanskrit, which is the ancestor of all Indo-Aryan languages. Sanskrit, which is a NOM-ACC language, as highlighted in (10a), used the instrumental marker *-ena* on the semantic agent within the passive construction, yet where the subject remains NOM-marked, as in (10b) (Kakati, 1941; Butt & Deo, 2001; Verbeke & De Cuypere, 2009).

[6] Although not discussed in the literature, one could argue that what we have in the case of (9), is an instance of the verb's agreement with the OBJ GF, rather than an instance of default agreement. This would in principle parallel Hindi in the sense that when ERG subjects are present in perfect contexts, the agreement which results on the verb is that with the object. While we won't engage in this discussion here, although such an analysis is a possibility, we refer the reader to (8a), at least if it can be said to constitute a like with like instantiation, and argue that if what we have in (9) were an instance of OBJ-verb agreement, rather than default 3rd PERSON agreement, then we would have expected to see 1st PERSON agreement on the verb in (8a), given the *-ok*-marked reflexive OBJ.

(10) a. devadatta-ḥ kaṭa-ṃ ca-kār-a
Devadatta-NOM mat-ACC PERF-make-3SG
'Devadatta made a mat.' (Verbeke & De Cuypere, 2009, p. 2)

b. devadatt-ena kaṭa-ḥ kṛ-taḥ
Devadatta-INS mat-NOM make-PST.PASS.PTCP
'The mat is made by Devadatta.' (Verbeke & De Cuypere, 2009, p. 3)

Although there are several accounts of how ergativity developed in Indo-Aryan languages, the reanalysis of a passive as an ergative construction is the most common hypothesis among scholars. Moreover, the evolution from the Sanskrit instrumental *-ena* or *-ī*, to the Assamese ergative marker *-e* seems highly probable (Kakati, 1941; Coghill, 2016; Kulikov, 2017). If we look at the timeline of this development, we find that Ashokan inscriptions from the Early Indo-Aryan (EIA) period show that the Sanskrit NOM marker *-ah* was being replaced by *-e*, as in *devānāmpiy-e* (*devānāmpri-yah*) 'the one who is loved by God' (Bloch, 1965; Devi, 1986).[7] Tagore (1948), as cited in Devi (1986, p. 68), proposed that the Sanskrit term *putrah* 'son' changed to *putte* in the Middle Indo-Aryan (MIA) period, until it eventually became *putti*, due to vowel weakening during the Apabharmśa period.

The Caryā texts composed by the Buddhists between the 8th and 12th century are claimed to bear the earliest evidence of literature stemming from the eastern group of Indo-Aryan languages. Devi (1986) notes several similarities between the Assamese ERG *-e*, and the *-e* and *i* subject markers found in these texts. Since the use of the *-e* marker had not become stable until the New Indo-Aryan (NIA) period, these texts bear the expected inconsistencies of the transition stage. For example, both *-e* and *i* were used, at this stage of the language, with the agent of transitive verbs, as in *sur-e* 'thief', *kānhi* 'Kānhā'.[8] However, towards the end of the texts, the use of *-e* gets stabilised as the sole subject marker. Devi (1986) points out that there is only one exclusive instance of an unmarked subject of a transitive verb in Caryā 6 of these texts. The example is represented in (11) below, where we observe an instance involving the subject NP *harina* 'deer' not taking an ERG marking, in spite of being the subject of two coordinated transitive clauses.

(11) tina na echupai hariṇa pibai na pāni
grass not touch deer drink no water
'The deer does not notch any grass nor does (the deer) drink any water.' (Devi, 1986, p. 70)

Apart from optionally-marked subjects of transitive verbs, there are also examples of optional *-e* marking on the subjects of intransitive verbs in these texts.

[7] Alternatively, as suggested by an internal reviewer this could potentially be a case of phonological change rather than (direct) replacement.

[8] Name of a Hindu God.

For example, in Caryā 48, the reflexive pronoun *apan*[9] 'self', which is the subject of the unaccusative verb *bah* 'sit' is marked with the *-e* marker, while the subject of the clause in the first conjunct: *grāhaka* 'customer' is unmarked, in line with the synchronic facts when in context of unaccusative intransitive verbs, such as *ai* 'come'.

(12) āilā grāhaka.∅ apan-e bahiā
come customer.NOM self-ERG sit.PST
'Customer came and (himself) sat down.' (Devi, 1986, p. 71)

This sporadic use of the *-e* marker on the subject of transitive verbs, and some intransitive verbs, can be taken as the stage where a split ergative system started emerging in Assamese. Assamese developed simultaneously with other eastern Indo-Aryan languages, such as Odia (Oriya) and Bengali from the common ancestor: Eastern Magadhi, which branched out of Māgadhi Prakrit in the MIA period (Chatterji, 1926). There is evidence that such NIA languages from the eastern branch, including Maithili also once used *-e* markers on their subjects (Chatterji, 1926; Kakati, 1941).

However, synchronically, Assamese differs from other eastern Indo-Aryan languages, including Bengali, Oriya, Maithili, and Bhojpuri, which have now lost their erstwhile ergative case system, and have become reanalysed as NOM-ACC systems. In contrast, Assamese, together with Sylheti, and Nepali, are the only eastern Indo-Aryan languages that have retained the ERG alignment of their parent language, but which is not based upon an ASPECT-based ergative split system.[10] Rather, they collectively display an intransitive-based split, which one could argue to be an influence akin to contact with neighbouring Tibeto-Burman languages. Devi (1986, p. 63) argues that the consistent use of *-e* that we see on agents in Assamese might be an influence from the Ahom (Tai) and Naga (Sino-Tibetan) languages, which mark their agents with a distinct marker. She further notes that texts from the 13th century that were composed just after the Ahoms conquered Assam show the use of an optional *-ko* marker with both NPs and pronouns. Moreover, the Nagas that were given place in the Ahom court used distinct agentive markers for their NPs and pronouns. The same can also be said for the Tangsa group of languages spoken to the east of Assam. Devi argues that the presence of such language systems in contact with Assamese must have accelerated the use, and later the consolidation of an agent marker in Assamese. A parallel can be drawn to Dakkhini (Stroński, 2010), which has lost its ERG case marking due to isolation from other Indo-Aryan languages, along with its long lasting influence from its neighbouring NOM-ACC Dravidian languages. Kakati (1941, p. 286), as mentioned earlier, on the other

[9]Note that Devi (1986) glosses the *-e* in this example as NOM. Here we gloss this morph as ERG. We additionally glossed the unmarked subject *grāhaka* 'customer' as a NOM and marked it with a ∅.

[10]Nepali *does* maintain an ASPECT-based split. However, this is only internal to the transitive sub-system (Li, 2007).

hand, argues that the ERG *-e* in Assamese is a reanalysis of the instrumental *-(er)e*[11] that is obligatorily present on the subjects of passive constructions built out of transitive verbs, as in: *hat-(er)e buwa kapur* 'cloth woven by hand'. He further argues that it is this constant use of the INS *-(er)e* that has lead to the habitual use of *-e* in the expression/realisation of agent subjects.

However, the synchronic analysis of the language shows that there is a distinction between the INS *-(er)e* and the ERG *-e*, even if the literature suggests that these were once the same *-(er)e* form in the past. Irrespective of the interchangeable use of the INS *-(er)e* and the ERG *-e*, it is the subjects with *-e* that render an agentive reading, and not the ones marked with *-(er)e*. The data in (13) is meant to demonstrate that although *kotari* 'knife' can be marked with *-e*, we still are glossing the morph as INS, as we cannot possibly assume two ERG-marked NPs in the clause. It is clear that in this active sentence, the ERG-marked 3.PL pronoun functions as the SUBJ.

(13) xĩhot=**e** tak kotari=re kat-il-e
3.PL=ERG 2.SG.ACC knife=INS cut-PST-3
'They cut him with a knife.'

Moreover, in sentences such as (14), *dak* 'post' can only be marked through the INS *-ere* marker. This suggests to us that an NP like *dak* 'post' can never be ascribed any agentive role, in contrast to the possibility with respect to *kotari* 'knife', which could be what is allowing us an *-e* morph to express the INS case in (13).

(14) sithi-k^{h}ɔn dak=ere/*=e ah-il
letter-CLF post=INS/=ERG come-PST
'The letter came by the post.'

2.3 Non-SUBJ case marking in Assamese

If we are to argue that DCM results in differential subject marking in Assamese, then we here present a context, where elsewhere in the grammar of the language we also observe distinct markings associated with the same GF. Here we consider the distribution of case in the context of non-SUBJ GFs. Just as it has been shown in the literature that the Animacy Hierarchy accounts for a good deal of the cross-linguistic variation in split ergative systems, with differences observed on the basis of the nature of the noun type (McGregor, 2009), the same premise can be applied to behaviours associated with OBJ GFs in Assamese, which come to be marked as ACC with the *-(o)k* marker.

The applicability of the Animacy Hierarchy scale may differ from language to another. It has, however been shown to have wider impact on a number of distinct

[11] In the early Assamese period, the INS *-(k)ere* was also used to express accompaniment, as in *jámāi-ere* 'with my son-in-law' (Kakati, 1941, p. 287). The Chittogong dialect takes both genitive *-ar* and instrumental *-di* on the same noun to express accompaniment as in *put-ar-di* 'with the son'(Kakati, 1941, pp. 286-287).

grammatical phenomena ranging from agreement to syntactic marking, and the like. Croft's (2003, p. 112) Animacy Hierarchy represented in (15), is indicative of the fact that, for instance, referents higher on the scale, such as 1st/2nd PERSON pronouns are more likely to receive overt case marking than inanimate common nouns lower on the hierarchy.

(15) **first, second-person pronoun < third-person pronoun < proper names < human common noun < nonhuman animate common noun < inanimate common noun**

While (animate-referring) pronouns in Assamese are always ACC-marked, as illustrated in (16), NPs do not display a uniform behaviour. For example, in (17a) the animate object *Rita* of the transitive verb *d^{h}or* 'hold' takes the ACC case marker *-(o)k*, while in (17b), the inanimate object *bol* 'ball', associated with the same transitive verb, remains unmarked. Leaving *Rita* unmarked in (17a), results in ungrammaticality. Such behaviours have been referred to as differential object marking (DOM) in the literature. DOM also exists in a number of typologically different languages, such as Turkish (Kornfilt, 2009), Maltese (Camilleri & Sadler, 2012), and Spanish (Comrie, 2013).

(16) a. xi.∅ muk dek^{h}-il-e
3.SG.M.NOM 1.SG.ACC see-PST-3
'He saw me.'

b. moi.∅ xihot=ok dek^{h}-il-u
1.SG.NOM 3.PL=ACC see-PST-1
'I saw them.'

(17) a. nitu=**e** rita=k/*rita d^{h}or-il-e
nitu=ERG rita=ACC/rita.∅ hold-PST-3
(Lit. 'Nitu held Rita.')
'Nitu caught Rita.'

b. nitu=**e** bol-tu/*bol-tu=k d^{h}or-il-e
nitu=ERG ball-CLF/*ball-CLF=ACC hold-PST-3
(Lit. 'Nitu held the ball.')
'Nitu caught the ball.'

Although non-human animates are higher on the Animacy Hierarchy than inanimates, in Assamese no distinction appears to be made between animate categories such as animals, birds, or trees and inanimates. This is illustrated through the data in (17b) and (18) that take no *-(o)k* ACC marking.

(18) a. goru-zoni.∅/*=k band^{h}-il-i-ne
cow-CLF.∅/=ACC tie-PST-1-Q
'Did you tie the cow?'

b. tamul-zupa.∅/*=k ne-kat-ib-i
areca-nut-CLF.∅/=ACC NEG-cut-FUT-2
'Do not cut the arica nut tree.'

As things stand, it seems therefore that DOM in Assamese is conditioned by a HUMAN feature. However, there is added intricacy to when and in which contexts does case marking appear even on non-HUMANS. For instance, if the *goru* 'cow' in (18a) is given a Proper Name, this will be *-(o)k* marked. If on the other hand, the Proper Name of an inanimate is in OBJ position, such as the *Taj Mahal* (a heritage monument), this will not get *-(o)k* marked (Chowdhary, 2014, p. 117). Beyond (ANIM) Proper Names (and pronouns), DOM on HUMAN NPs is interrelated with concerns that pertain to SPECIFICITY. This is infact something that has been discussed quite amply for Hindi (see e.g. Butt (1993) and Montaut (2018)).

Assamese is a numeral classifier language. This implies that once a classifier attaches onto the right-edge of an NP[12] OBJ (be it HUMAN or non-HUMAN) in the absence of a numeral, which would otherwise take the classifier, that NP becomes DEFINITE. While the behaviour of DEF/INDEF cuts across the board irrespective of whether a HUMAN or non-HUMAN OBJ is involved, as illustrated through (19) and (20 a-b) below, the addition of case becomes obligatory in the context of a [+SPEC] reading in association with HUMAN NPs, as illustrated in (20 c).

(19) a. moi.∅ kitap.∅/*=ok porh-i b^{h}al pa-o
1.SG.NOM book.∅/=ACC read-PROG good get.1
(Lit: 'I feel good reading book.')
'I love reading books.' ([+/- DEF] [- SPEC])

b. moi.∅ kitap-k^{h}ɔn/*=ok porh-i b^{h}al pa-o
1.SG.NOM book-CLF/=ACC read-PROG good get.1
'I love reading the book.' ([+ DEF] [+/- SPEC])

(20) a. pulis=**e** sur.∅ d^{h}or-e
police=ERG thief.∅ hold.PRES-3
(Lit: 'Police hold thief.')
'Police catches thieves.' ([+/- DEF] [- SPEC])

b. pulis=**e** sur-tu.∅ d^{h}or-il-e
police=ERG thief-CLF.∅ hold-PST-3
'Police caught the thief.' ([+ DEF] [+/- SPEC])

[12]Note that when there is a classifier as well as a case marker attached onto an NP, the classifier always precedes the case marker.

c. pulis=**e** sur-tu=k d^hor-il-e
police=ERG thief-CLF=ACC hold-PST-3
'Police caught the thief.' ([+ DEF] [+ SPEC])

What this implies therefore is that HUMAN NPs are made SPECIFIC via the very presence of *-(o)k* marking; a strategy which is not morphosyntactically available for non-HUMAN NPs. In contrast, SPECIFICTY in non-HUMAN INANIM NPs such as *kitap* 'book' comes solely from the context. Furthermore, although the majority of the literature (e.g. Gundel et al. (1993), Enç (1991)) suggest that DEFINITENESS also implies SPECIFICITY, this does not hold true for Assamese, given that the numeral classifier *-tu* attached to the HUMAN NP *sur* 'thief' implies that it already takes a DEFINITE reference, but which is not yet made SPECIFIC, necessarily, until a case marker is present on OBJ. While DOM brings out SPECIFICITY effects in both Hindi and Assamese, with Assamese allowing this only in the context of HUMAN NP OBJs, there are other Indo-Aryan languages like Sinhala/Sinhalese, where an ANIMACY is all that matters in the determination of whether objects can be optionally ACC-marked or not (Thampoe, 2017).

It should finally be noted here that DOM in Assamese only applies to objects in neutral contexts. If the object is placed in a non-neutral context, such as in a topicalised position, typically left-adjacent to the verb (along with additional intonation cues) the inanimate indefinite NP must be ACC-marked. Such a behaviour is highlighted in (21) below through the inanimate, indefinite NPs *zibon* 'life' and *d^hopāt* 'tobacco'.

(21) zibɔn-ok adɔr-ok d^hopat-ok nɔ-hɔi
life-ACC welcome-IMP tobacco-ACC NEG-be.PRES
'Welcome life, not tobacco' (Chowdhary, 2014, p. 118)

The morphosyntax and the structure of (21) would imply that a structure such as (22), although displaying a parallel string, cannot be understood as a topicalisation structure. What we have in (22) is an instance where the INANIM *pani* 'water' and *mod* 'alcohol' are unmarked, unlike the ACC marking on the topicalised *zibon* 'life' and *d^hopat* 'tobacco'. The non-marking of the NPs in (22) is in line with them being INANIM NPs sitting low on the Animacy Hierarchy. For this reason therefore, as also suggested by an internal reviewer, what we have here is a case of an SOV structure with the SUBJ dropped by virtue of the imperative mood of the structure.

(22) pani.∅ kha-ok mod.∅ nɔ-hɔi
water.NOM drink-IMP alcohol-NOM NEG-be.PRES
'Drink water not alcohol.'

So far we have only considered what goes on with primary/direct objects, or OBJ GFs in LFG terms. When we turn our attention to indirect objects, i.e. those

GFs that function as recipients in ditransitive constructions, we have evidence, although not given any attention in the literature on Assamese that morphologically, the OBJ$_{\theta}$ takes a distinct case distribution, even if, the marker which we here, for expository convenience refer to as DAT, takes a homophonous *-(o)k* form just as the ACC, (as is also the case in Hindi/Urdu). What is key for us, in the light of the data paradigm in (23), is that the recipient, which can be a Proper Name, as in (23a), a HUMAN NP as in (23b), a non-HUMAN ANIM NP as in (23c), and an INANIM NP as in (23d) is that of a double object construction. Evidence that the recipient in Assamese maps onto an OBJ$_{\theta}$, rather than onto an OBJ, is clear from the distribution of *-(o)k* ACC marking on the theme, which patterns exactly what we have just discussed above. On the other hand, *-(o)k* as a DAT marker on the OBJ$_{\theta}$ does not display a similar behaviour. Rather, such marking is present throughout. Note that in Assamese a clear constituent order preference holds whereby it is more likely to have the recipient argument preceding the theme.

(23) a. tai.∅ pinki=k/*∅ puna=k/*∅ hop-il-e
3.SG.NOM Pinki=DAT/∅ Puna=ACC/∅ entrust-PST-3
'She entrusted (the custody of) Puna to Pinki.' (Adapted from (Chowdhary, 2014, p. 119)

b. tai.∅ mastor=ok/*∅ lora-tu(=k) hop-il-e
3.SG.NOM teacher=DAT/∅ boy-CLF(=ACC) entrust-PST-3
'She entrusted (the custody of) the boy to the teacher.' (Adapted from (Chowdhary, 2014, p. 119)

c. tɛõluk=**e** kukur-tu=k//*∅ b^{h}at/*∅ d-il-e
3.PL=ERG dog-CLF=ACC/∅ rice/*∅ give-PST-3
'They gave rice to the dog.'

d. tɛõluk=**e** xoŋgram-tu=k/*∅ notun ort^{h}o/*∅ d-il-e
3.PL=ERG revolution-CLF=ACC/∅ new meaning/∅ give-PST-3
'They gave new meaning to the revolution.'

We here, finally, consider prepositional objects that are GEN-marked via the phonologically-conditioned allomorphs *-(o)r*, as shown by the oblique objects *Pinki* in (24a), *deutak* 'father' in (24b), and *duwar* 'door' in (24c) .

(24) a. razu=**e** pinki=r karone p^{h}ul kin-is-e
razu=ERG pinki=GEN for flower buy-PERF-3
'Razu has bought flowers for Pinki.'

b. tai.∅ deutak=or karone sit^{h}i lik^{h}-is-e
3.F.NOM father=GEN for letter write-PERF-3
'She has written a letter for her father.'

c. tai.∅ duwar-k^{h}ɔn=or karone tola e-ta kin-il-e
3.F.NOM door-CLF=GEN for lock one-CLF kin-PST-3
'She bought a lock for the door.'

What this means for us is that OBL OBJs, i.e. the OBJ GFs which Ps subcategorise for, are special. Beyond the fact that they get GEN-marked, such GEN marking appears to 'block' a distribution that parallels what we have described above in the context of ACC-marked OBJs of Vs.

Table (2) below now provides a characterisation of the distribution of case marking across the non-SUBJ GFs.

Value	OBJs	OBJ_{θ}	OBL OBJs
Pronoun	ACC	DAT	GEN
Proper Names	ACC	DAT	GEN
HUMAN NP	(ACC)	DAT	GEN
ANIM/INANIM NPs	∅	DAT	GEN

Table 2: Distribution of case-marking on non-SUBJ GFs

From the above discussion it transpires that the observations associated with the OBJ GF, in relation to the distribution of case correlates with SPECIFICITY. One could say that the behaviours attributed to the Animacy Hierarchy fall out in an expected manner, since for instance personal pronouns and Proper Names are inherently specific, and hence precisely illustrate the contexts where we get to observe obligatory ACC marking. In contrast, OBJ_{θ} and OBL OBJs display a uniform behaviour and do not pertain to any Animacy Hierarchy-based observations.

2.4 Current predictions of change in progress

Focusing specifically on the SUBJ GF of intransitive verbs in Assamese, and the distribution of ERG case, it seems to us that change is in progress. The hypothesis of the change we envisage can be summarised as follows. The ERG-based split in intransitives does no longer seem to be *solely* motivated by agentivity, but rather, it has started infiltrating within the unaccusative domain, and wherein, it is being guided by a distinct semantic factor, namely ANIMACY. In support of this hypothesis is the child language data of Saikia (in prep.), as well as a young adult based study which was designed to function as a control group, but ended up interestingly displaying parallel results. A discussion of the study and its results follows below.

3 The study

3.1 Methodology

The research on which this study is based, is part of a larger study that aims to look at children's acquisition of split ergativity in Assamese (Saikia, in prep.). All the data for this study was collected from the districts of Tinsukia and Dibrugarh in eastern Assam. For the first part of this study, 40 children (2-6 years) whose primary language input is Assamese took part in a Contrastive Elicitation Task for Testing Case Marking (Ruigendijk, 2015). Further, to develop a set of comparable data, 22 Assamese speaking adults were asked to take part in the same elicitation task experiment. The participants in this control group were young adults (16-25 years) studying at a higher educational institute. All the participants of this production task were asked to describe 12 pairs of minimally contrastive images, some of which are shown in Figure 1.[13]

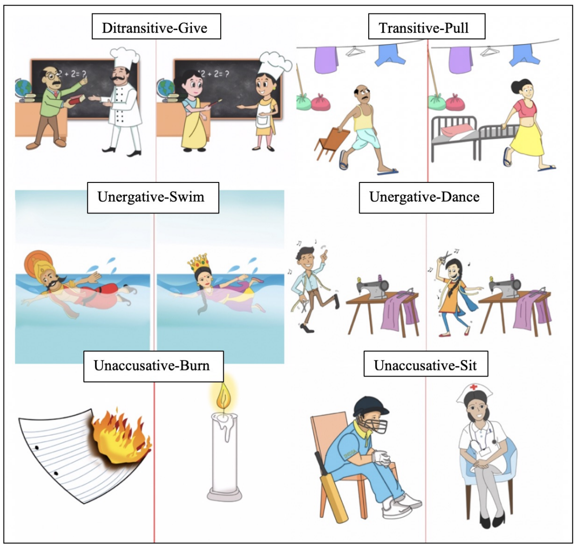


Figure 1: Contrastive Elicitation Task for Testing Case Marking (based on Ruigendijk (2015))

Ruigendijk (2015) Contrastive Elicitation Task was originally designed for two verb conditions: ditransitive and transitive. However, since what Saikia (in prep.) is after, which encompasses the whole case alignment in Assamese, and with the knowledge that Assamese has an intransitivity-based split, intransitives, specially four unergative, and four unaccusative verbs were included in the stimuli, and a new set of pictures were designed to suit any Indian language and culture. The

[13]The entire task involved describing the illustrations of twelve different verbs in both progressive and perfect structures. However, notwithstanding the incorporation of this grammatical ASPECTual distinction, no correlation was observed with respect to SUBJ case marking, and hence we do not discuss it further.

stimuli were controlled for conditions such as verb type, PERSON, NUMBER and ANIMACY to elicit target utterances for the specific case markers. The intransitive verb types, in particular, included the unergative *nas* 'dance', *xãtur* 'swim', *zopia* 'jump', and 'dour' run, and the unaccusative *por* 'fall', *zol* 'burn', *boh* 'sit', and *dub* 'sink'. Out of the set of these eight verbs, only the subjects of 'fall' and 'sit' were HUMAN. ANIM/HUMAN subjects for 'burn' and 'sink' were avoided given the projected violent nature, as majority of our participants were small children. All the unergative verbs in the study involved HUMAN subjects. We here deem important to reiterate why the task did not include any stimuli that involved unergative verbs with INANIM subjects in the elicitation task. This is because, as illustrated clearly in §2.1, through the pair in (2), ANIMACY plays no role in the assignment of ERG case marking. Moreover, as the hypothesis posited in §2.4 already mentions, the observed change is exclusively taking place in the domain of unaccusative verbs.

3.2 Results and discussion

All the participants of the elicitation task were observed to be adhering to the description of the transitive and ditransitive structures as provided in §2.3. However, we got a mixed response in the context of intransitive verbs. Since the stimuli were developed following the description in §2.1, we were expecting that the subjects of unaccusative verbs will be ∅, i.e. NOM, while the subjects of unergative verbs will maintain their *-e* ERG marking. However, our data did not reflect such a clear agentivity-based intransitive split. In fact, we found that both children and adult participants alike were ignoring the intransitivity split discussed as described earlier, and were rather embracing a new case marking pattern, which appears to be conditioned by a distinct semantic factor of the nominal.

The main evidence for this observed behaviour comes from the infiltration of the erstwhile ERG morph, as a marker of a [ANIMATE +] feature-value in the f-structure of the S_o SUBJ of unaccusative PREDs. This is in contrast to its previous canonical function as a marker of the S_a SUBJ of unergative PREDs, as a means with which to exhibit their agent thematic role.

The observed change is happening at an average of 20% of the time in the adult data, and 73% of the time in the child language data. We take this to be possibly demonstrating the impetus of the change in place, and its direction. This emergent ANIMACY-based split in the SUBJ system of unaccusatives is interestingly yet another semantic factor over and above the semantic factors that condition the splits that guide DOM in Assamese as discussed in §2.3.

The data in hand supporting this observation comes from the contrast presented in (25) vs. (26). Here we have the unaccusative predicates *boh* 'sit' and *por* 'fall' taking ANIM subjects where we observe the emergent *-e* ERG marking as opposed to the predicates *zol* 'burn' and *dub* 'sink' with INANIM subjects which in turn remain ∅-marked. In fact, all the participants consistently maintained a ∅ marking in such instances.

(25) a. k^h^eluwoi-zon=**e**/narse-goraki=**e** boh-i as-e
sportsman-CLF=ERG/nurse-CLF=ERG sit-PROG be.PRES-3
'The sportsman/nurse is sitting.'

b. bimansalok-zon=**e**/bimansalika-goraki=**e** por-i as-e
pilot.M-CLF=ERG/pilot.F-CLF=ERG fall-PROG be.PRES-3
'The pilot (M/F) IS FALLING.'
(**Emergent** ANIM ERG SUBJ **of unaccusative** PRED)

(26) a. kagos-k^{h}ɔn.∅/mom-dal.∅ zol-i as-e
paper-CLF.NOM/candle-CLF.NOM burn-PROG be.PRES-3
'The paper/candle is burning.'

b. nao-k^{h}ɔn.∅/bakos-tu dub-i as-e
boat-CLF.NOM/box-CLF.NOM sink-PROG be.PRES-3
'The boat/box is sinking.'
(INANIM NOM SUBJ **of unaccusative** PRED)

The same pattern was also noticed in the case of unergative verbs, which should otherwise, in accordance with the ERG split system, take *-e* marked SUBJ.

We hypothesise this new emerging situation to have arisen as a result of a reanalysis of what the morphological form that is responsible for the unaccusative-unergative split, i.e. the *-e* that exists in the intransitive domain, comes to express. The ERG's erstwhile agentive marking has, within the unaccusative domain of intransitives seemingly come to express an ANIMACY distinction. Consequently, the split is being overhauled, in the sense that it is now being conditioned by a semantic feature in the lexical entry, rather than by a theta-role - GF association at the argument-structure.

The emerging system is represented in Table (2).

Value	A	S_a	S_o
Animate	-e	-e	**-e**
Inanimate	-e	**-**∅	-∅

Table 3: **The emergent** ANIMACY**-based split**

4 Conclusion

The case alignment system in Assamese is currently undergoing change, and a shift appears to be taking place when SUBJs are marked. A representation of this illustrated in Figure (2).

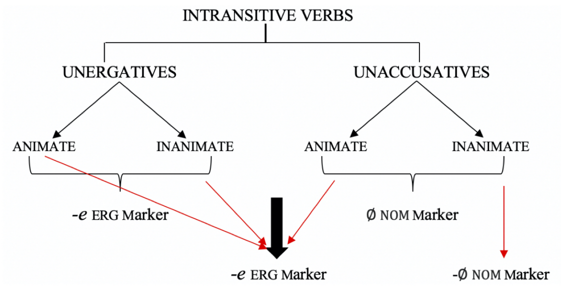


Figure 2: Current state and the change in progress within intransitives

While the split in ERG case on SUBJs (excluding any topicalisation or emphatic effects) depended exclusively on the thematic role of the SUBJ, as influenced by the nature of the verb, i.e. depending on whether it is unergative or unaccusative, synchronically a split is emerging. Internal to the unaccusative domain of the intransitive predicates there appears to be a split dependent on ANIMACY. It remains to be seen however, whether there may be any potential effects coming from the lexical aspect of the different intransitive verbs. Moreover, perhaps as a self criticism associated with the design of the study, a flaw is noticeable which could impinge on our findings. Since the ANIMATE SUBJs of the unaccusative predicates used in the study happened to be all HUMAN, the emerging distinct use of the ERG marker in the domain of the unaccusatives may well be more fine-grained. It may be one which solely considers HUMAN vs. non-HUMAN SUBJs, rather than a broader ANIMATE vs. INANIMATE distinction.

References

Ahmed, Tafseer. 2010. The unaccusativity/unergativity distinction in Urdu. *Journal of South Asian Linguistics* 3(1).

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. *Natural Language & Linguistic Theory* 17(4). 673–711.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483.

Amritavalli, R. & Partha P. Sarma. 2002. A case distinction between unaccusative and unergative subjects in Assamese. *Snippets* 5. 6–7.

Bloch, Jules. 1965. *Indo-Aryan: From the Vedas to modern times*. Paris: Librarie d'Amérique et d'Orient Adrien-Maisonneuve.

Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. In *29th Regional Meeting of the Chicago Linguistic Society*, vol. 1, 80–103.

Butt, Miriam. 2006. The dative-ergative connection. *Empirical Issues in Syntax and Semantics* 6. 69–92.

Butt, Miriam & Tafseer Ahmed. 2011. The redevelopment of Indo-Aryan case systems from a lexical semantic perspective. *Morphology* 21(3-4). 545–572.

Butt, Miriam & Ashwini Deo. 2001. Ergativity in Indo-Aryan. *Kurdica* 5(3).

Butt, Miriam & Tracy Holloway King. 1991. Semantic case in Urdu. In *27th Regional Meeting of the Chicago Linguistic Society*, vol. 27 1, 31–45.

Camilleri, Maris & Louisa Sadler. 2012. On the analysis of non-selected datives in Maltese. In Miriam Butt & Tracy Halloway King (eds.), *Proceedings of the LFG12 Conference*, 118–138. Stanford: CSLI Publications (online publications).

Chatterji, Suniti K. 1926. *The origin and development of the Bengali language*. Calcutta: Calcutta University Press.

Chowdhary, Runima. 2014. Differential marking of cases in Asamiya. *North East Indian Linguistics* 6. 107–124.

Coghill, Eleanor. 2016. *The rise and fall of ergativity in Aramaic: Cycles of alignment change*, vol. 21. Oxford: Oxford University Press.

Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Texas Press.

Comrie, Bernard. 2013. Alignment of case marking of full noun phrases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/98.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Devi, Jayantimala. 1986. *Ergativity: A Historical Analysis in Assamese*. New Delhi: University of Delhi dissertation.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55. 59–138.

Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22. 1–25.

Goswami, Golok C. & Jyotiprakash Tamuli. 2003. Asamiya. In George Cardona & Dhanesh Jain (eds.), *The Indo-Aryan Languages*, 391–443. London: Routledge.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.

Haddad, Youssef A. 2011. *Control into conjunctive Participle Clauses: The case of Assamese*. Berlin: De Gruyter Moutone.

Handschuh, Corinna. 2008. Split Marked-S Case Systems. *Scales* 86. 73–82.

Kakati, Banikanta. 1941. *Assamese: Its Formantion and Development*. Gauhati, Assam: Department of Historical and Antiquarian Studies.

Kornfilt, Jaklin. 2009. DOM and two types of DSM in Turkish. In de Swart P. de Hoop H. (ed.), *Differential subject marking*, 79–111. Dordrecht: Springer.

Kulikov, Leonid. 2017. Indo-aryan. In Mate Kapović (ed.), *The Indo-European Languages*, 214–262. London: Routledge.

Li, Chao. 2007. Split ergativity and split intransitivity in Nepali. *Lingua* 117(8). 1462 – 1482.

Lightfoot, David. 2010. Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5). 677–684.

McGregor, William B. 2009. Typology of ergativity. *Language and Linguistics Compass* 3(1). 480–508.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. Stanford: CSLI Publications.

Montaut, Annie. 2018. The rise of differential object marking in Hindi and related languages. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, vol. 19, 281–313. Berlin: Language Science Press.

Nath, Diganta. 2003. *Case in Assamese*. Hyderabad: Central Institute of English and Foreign Languages dissertation.

Ruigendijk, Esther. 2015. Contrastive elicitation task for testing case marking. In Sharon Armon-Lotem, Jan de Jong & Natalia Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment*, 38–54. Bristol: Multilingual Matters.

Saha, Atanu & Bipasha Patgiri. 2013. Ergativity in Axomiya. *Language in India* 13(12). 35–46.

Saikia, Pori. in prep. *Children's Acquisition of Assamese Split Ergativity*. Colchester: University of Essex dissertation.

Stroński, Krzysztof. 2010. Variation of ergativity patterns in Indo-Aryan. *Poznań Studies in Contemporary Linguistics* 46(2). 237–253.

Thampoe, Harold D. 2017. *Sinhala and Tamil: A case of contact-induced restructuring*. Newcastle: Newcastle University dissertation.

van Valin Jr, Robert D. 1990. Semantic Parameters of Split Intransitivity. *Language* 66. 221–260.

Verbeke, Saartje & Ludovic De Cuypere. 2009. The rise of ergativity in Hindi: Assessing the role of grammaticalization. *Folia Linguistica Historica* 30. 1–24.

Zakharyin, Boris. 2015. Indo-Aryan ergativity and its analogues in languages of central and western eurasia. *Lingua Posnaniensis* 57(2). 63–76.

Computational Challenges with Tamil Complex Predicates

Kengatharaiyer Sarveswaran
University of Moratuwa

Miriam Butt
University of Konstanz

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 272–292

http://csli-publications.stanford.edu/LFG/2019

Keywords: complex predicates, FSM, Tamil, restriction operator, morphology-syntax interface

Sarveswaran, Kengatharaiyer, & Butt, Miriam. 2019. Computational Challenges with Tamil Complex Predicates. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 272–292. Stanford, CA: CSLI Publications.

Abstract

This paper presents work in the context of the development of a computational ParGram style grammar for Tamil. The grammar is implemented via the XLE grammar development platform and contains a Finite-State Morphological analyser implemented via Foma. This paper reports on challenges for the implementation found with respect to V-V complex predicates in terms of the interaction with phonology (*Sandhi*) and the lexicon. In particular, we focused on the interaction of causation and passivisation with complex predication. This paper provides further evidence from Tamil complex predicates for the use of the Restriction Operator and also addresses issues with respect to complex predication at the morphology-syntax interface.

1 Introduction

This paper presents work in the context of the development of a computational ParGram (Butt et al. 1999) style grammar for Tamil.[1] The grammar is implemented via the XLE grammar development platform (Crouch et al. 2017) and contains a finite-state morphological (FSM) analyser implemented (Sarveswaran et al. 2019) via Foma (Hulden 2009). The work to date has mainly focused on the implementation of basic clause types and the inflectional morphology within the morphological analyser.

In pursuing this work, we encountered challenges with respect to the implementation of V-V complex predicates in terms of the interaction with phonology, the lexicon and derivational morphology. In this paper, we focus on the challenges arising with respect to the interaction of causation and passivisation within complex predicates. Similar but not identical issues have been noted for Turkish (Çetinoğlu 2009) and Urdu (Bögel et al. 2019), leading to the use of the Restriction Operator for passivisation, rather than the classical lexical rules of LFG. This paper provides further evidence for the use of the Restriction Operator from Tamil complex predicates and also addresses issues with respect to complex predication at the morphology-syntax interface that have not previously been encountered within ParGram.

Tamil is well known for its diverse types of V-V sequences (Steever 1987, 2005). Here we focus on an instance of V-V complex predication as discussed by Annamalai (2013). We illustrate how this type of complex predication is handled in the Tamil LFG grammar using the causative and passive constructions of two verbs: ‘buy’ and ‘give’, whereby ‘give’ functions as a light verb that adds a beneficiary to the overall predication. A particular challenge in Tamil is that the elements of complex predicates can either be found written together as a single word, or be separated into two tokens. However, phonological Sandhi phenomena apply irrespective of the expression

[1] We gratefully acknowledge funding from the DAAD (German Academic Exchange Office) in support of this research.

in terms of one or two tokens and are realised obligatorily within Tamil orthography. The phonological properties of one part of the complex predicate condition Sandhi rules on the other part, irrespective of whether these are written as one or two parts. While this points towards an overall realisation of one prosodic unit irrespective of the realisation in terms of one vs. two tokens, it poses a challenge for the computational implementation of morphology-syntax interface as the analysis of individual words within the morphological analyser must anticipate possible Sandhi rules triggered by complex predicate formation in the syntax. We show how this phenomena can be handled without an extension of the existing ParGram architecture.

2 Background

2.1 Tamil

Tamil is a Southern Dravidian language spoken natively by more than 80 million people across the world. It has been recognised as a classical language by the government of India since it has more than 2000 years of a continuous and unbroken literary tradition (Hart 2000). It is an official language of Sri Lanka and Singapore, and has regional official status in Tamil Nadu and Pondichchery, India.

Tamil words have been primarily divided into four types, namely: nouns, verbs, intensifiers/attributives, and particles in grammar books written by native grammarians (Thesikar 1957, Senavaraiyar 1938). However, more modern work provides a different type of classification (Nuhman 1999, Paramasivam 2011). Beyond the nature of their part-of-speech category, words in Tamil can be further classified into divisible and indivisible categories. A divisible word can have six parts, namely: root, suffix, medial particle, *chariyai*, *Sandhi* and alteration (Nuhman 1999, Senavaraiyar 1938), where medial particles can be tense markers, and *chariyai* is a phonological modifier which can be further divided into a euphonic marker and an oblique marker based on the function expressed by it (Lehmann 1993). The notion of *Sandhi* is elaborated upon in the next section. The alteration is a phonological change which is realised as such in the orthography.

(1)

வந்தனன் (vantanan)

வா	த்(ந்)	த்	அன்	அன்
vaa	t(n)	t	an	an
root (வா-> வ)	*Sandhi* (த் -> ந்)	*medial*	*chariyai*	suffix

'(He) came.'

Example (1) shows that how a divisible word can be sliced into different parts. However, not all the divisible words have all these six parts.[2] In (1), வா-> வ and த் -> ந் are called alterations.

2.2 சந்தி (Sandhi)

Internal *Sandhi* refers to a phonological process triggered across two morphs within a (prosodic) word. When such a process is applied at the boundary of two words it is referred to as external *Sandhi*. External Sandhi can occur when the second word begins with one of the following consonants: க் (k), ச்(c), த் (t), ப் (p). However, further licensing conditions also need to be met, as shown below. Internal *Sandhi* is purely morphophonological in nature, while external *Sandhi* is also subject to syntactic or semantic constraints. Example (2) shows an internal *Sandhi* [t], this is inserted because the past tense marker (t) follows a vowel. Since Tamil orthography closely reflects the phonology of the language, *Sandhi's* effects on the orthography must necessarily be dealt with by any Tamil computational grammar.

(2)

படித்தான் (padittaan)

படி	-த்	-த்	-ஆன்
padi	-t	-t	-aan
study	-SAN	-PAST	-3SMR

'(He) studied.'

The examples in (3) and (4) illustrate a case of external *Sandhi*. The object ('bull') and the verb contain identical final (object) and initial (verb) phonological segments. However, in (3) the insertion of *Sandhi* [p] is obligatory: *Sandhi* must apply if there is an overt accusative on the object. However, as shown in (4), no *Sandhi* occurs when there is no accusative marker even though it is an equivalent construction in terms of segmental phonology, i.e. in both (3) and (4) /i/ is the final vowel in the noun preceding the verb பிடித்தான் (pidiththan).

(3)

கந்தன்	காளையைப்	பிடித்தான்
kanthan	kalai-yai-p	pidiththan
Kanthan.NOM	bull-ACC-SAN	catch.PAST.3SMR

'Kanthan caught the bull.'

[2]Abbreviations in the glosses are: VP=Verbal Participle; INF=Infinitive; 3SN=3rd Person Singular Neuter; 1S=1st Person, Singular; 3SMR=3rd Person, Singular, Masculine and Rational; PASS=Passive; SAN=Sandhi; RP= Relative Participle; IMP=Imperative; CAUS=Causative; NOM=Nominative; DAT=Dative; ACC=Accusative.

While having *Sandhi* in (3) is compulsory, including the *Sandhi* in (4) is considered ungrammatical. This thus illustrates that *Sandhi* is not conditioned by purely segmental phonological factors.

(4)

கந்தன்	ஒரு	காளை	பிடித்தான்
kanthan	oru	kalai	pidiththan
Kanthan.NOM	a	bull.NOM	catch.PAST.3SMR

‘Kanthan caught a bull.’

The presence or absence of *Sandhi* is furthermore indicative of different underlying syntactic structures. For instance, as shown in (5)–(6), the *Sandhi* [c] surfaces when the token is an adjective in (5), but does not surface in (6), where a string identical item is functioning as a relative participle and is at the right edge of a relative clause boundary. In (5) the item is the adjective முக்கியம் (mukkiyam) ‘important’, in (6) it is a relative participle derived from the verb முக்கு (mukku) ‘dip’.

(5)

அவன்	முக்கியச்	செய்திகளை	வாசித்தான்
avan	mukkuya-c	seithikal-ai	vasiththan
avan.NOM	important.ADJ-SAN	news.PL-ACC	read.PAST.3SMR

‘He read the important news items.’

(6)

அவன்	முக்கிய	செய்திகளை	வாசித்தான்
avan	mukkuya	seithikal-ai	vasiththan
avan.NOM	dip.RP	news.PL-ACC	read.PAST.3SMR

‘He read dipped news items.’

We assume that the presence or absence of *Sandhi* is related to whether items are phrased together prosodically or not, so that *Sandhi* occurs within a prosodic phrase, but not across prosodic boundaries (see also Lahiri & Fitzpatrick-Cole 1999 for Bengali). Dealing with such prosodically conditioned *Sandhi* provides a challenge for computational grammar development within the ParGram framework. In this paper we show how we can model the phenomena via an interaction of the FSM analyser with the computational syntax.

2.3 Verbs in Tamil

Verbal morphology on simplex verbs in Tamil expresses information about tense, mood, aspect, negation, interrogativity, emphasis, speaker perspective, sentience or rationality, and conditional and causal relations (Annamalai et al. 2014). The structure of a simple verb is <root> + <medial-particle> + <terminal-suffix>. However, there are cases where a euphonic particle is also added in the middle, in addition to the usual medial particle. The medial particle is mainly used to realise the tense of the verb. There are three values for tenses in Tamil: past, present and future (Pope 1979, Lehmann 1993, Paramasivam 2011).

The terminal-suffix of a finite verb is used to realise multiple types of information such as number, person, gender, and rationality (Pope 1979) or status (Lehmann 1993). As for other morphosyntactic features, Tamil has singular and plural as values for number, 1st/2nd/3rd person values, and three gender values, namely masculine, feminine and neuter. In addition to these three genders, epicene is also used as a fourth one to mark the 3rd person plural forms (Lehmann 1993). Entities in Tamil are fundamentally classified into rational or irrational. This split is based on the status of an entity: Entities are termed rational if they are perceived as being able to think on their own, whereas the rest are termed irrational. This is different from splits found otherwise in terms of human vs. non-human or animacy. For instance, infants are considered to be irrational like other animal or inanimate objects even though infants are human and animate. Further, when people behave as if they are insane, they are morphologically classified as irrational.

2.4 ParGram project

The Parallel Grammar (ParGram) Project (Butt et al. 1999, Butt & King 2002) aims to develop and implement large and wide coverage grammars for languages of different families. These parallel grammars are written collaboratively within the linguistic framework of LFG and with an agreed set of grammatical features by the project group members. The XLE (Xerox Linguistic Environment) (Crouch et al. 2017), which is a parsing and generation implementation of LFG provided by PARC, is used as a grammar development platform. In addition to putting effort into feature standardisation, the project also promotes similar analyses for similar phenomena across languages (Butt & King 2002), a property which is useful for crosslingual language applications like machine translation and information retrieval.

3 Complex Predicates

The study of complex predicates (CP) has received a great deal of attention in the linguistic literature, along with a number of distinct interpretations. We base this paper on the definition proposed by Butt (1995), which views CPs as being formed when two or more predicational units enter into a relationship of co-predication. Each predicational unit adds arguments to a mono-clausal predication; a similar definition or idea can also be found in

Mohanan (1994, 1997) and Alsina et al. (1997). In LFG, these two or more semantic heads correspond to a single PRED at the level of f-structure. C-structure does not determine CP status and the elements contributing a CP can be either morphological or syntactic (Butt 2010). However, regardless of whether the complex predication is morphological or syntactic, the composition of the arguments of both of the predicational units works according to the same principles (Alsina 1996).

Complex predicates are very common in Tamil (Annamalai 2013). For instance, verbs like வை (vay) 'place', விடு (vidu) 'let go', பார் (paar) 'see/look' may function as both main/full and light verbs. As light verbs, they mean 'cause', 'let' and 'try', respectively (Annamalai 2013).

3.1 Complex Predicates in Tamil

Tamil verbs have been analysed mostly from a prescriptive perspective, and most of these studies are based on the very first Tamil grammar called *Tholkapiyam*[3] and a derived piece of work, the *Nannool*, published in the 13th century CE. From the 18th century CE on Western scholars have also contributed to the study of Tamil grammar. However, except for the attempt by Annamalai (2013), no scholars have clearly articulated differences between complex predicates, serial verb constructions (Steever 2005, Fedson 1981), complex verbs (Agesthialingom 1971), and compound verbs (Agesthialingom 1971, Nuhman 1999, Fedson 1981, Paramasivam 2011). This stands in contrast to the work done for other South Asian languages like Urdu/Hindi (Butt 1995, Mohanan 1997, Butt & Lahiri 2013). However, it is important to understand the differences across these potentially confusing categories for the development of computational resources such as computational grammars (Butt & King 2002), WordNet (Chakrabarti et al. 2007), and machine translation (Kaplan & Wedekind 1993, Butt 1994).

As noted in the existing literature (Annamalai 2013, Steever 2005), Tamil is well known for diverse types of Verb-Verb (V-V) and Noun-Verb (N-V) constructions. Muthuchchanmugan (2005) shows that the main verb (also called lexical head word) in Tamil can be followed by up to four verbal units. However, whether all of them are auxiliaries as he claims, or not, is debatable. Tamil is a head-final language. In V-V constructions the terminal verbal unit is the final item in a sequence. The preceding verbal units can be in either an adverbial or infinitival form. The terminal verbal unit is the item that carries all the functional information such as tense, person, number, and gender. The V-V sequences are used to express a range of semantic information. This includes crosslinguistically well-established categories such as causative, passive, permissive, negation, aspectual information, and mood and modality, including obligation vs. possibility. The

[3]The date of publication is imprecise and uncertain, scholars argue that it could be between the 5th century BCE and the 5th century CE.

literature also describes definitive and conclusive meanings, the expression of irritation, carelessness, augmentation, prediction and intention (Paramasivam 2011, Muthuchchanmugan 2005). Tamil also has compound nominal predicates with N-N and V-N sequences, but these are not the topic of this paper. In what follows, we focus on the treatment of light verbs, causatives and passives in V-V constructions.

3.2 Light Verb Constructions

Light Verbs (LV) differ from main/full verbs in terms of their syntactic distribution and lexical semantics. While main verbs can stand alone and predicate independently, light verbs are dependent on the existence of another predicative element in the clause. LVs are light in the sense that they do not carry the meaning of the corresponding full verb, yet they still contain lexical semantic information (Butt 2010, Annamalai 2013). Unlike auxiliaries, they are not fully functional elements. Together with the main predicational element, in our case a verb, the light verb forms a syntactically monoclausal unit. Following (Butt 2010), we analyse LVs as a separate syntactic category and differentiate them from both main verbs and auxiliaries in the language. Such structures are common in South Asian Languages (SAL) (Butt & Lahiri 2013), including Tamil (Annamalai 2013).

Annamalai (2013) has analysed various V-V, Infinitive-V, N-V and Verbal Participle-V sequences and differentiates between Serial Verb Constructions (SVC) and Complex Predicates (CP). Example (7) illustrates a simple transitive verb வாங்கு (vangu) 'buy'. The same main verb used together with கொடு (kodu) 'give' in its light verb sense forms a CP in (8). The light verb 'give' contributes a beneficiary meaning to the predication and licenses the use of an additional beneficiary indirect object (OBJ-TH). The light verb as the terminal verbal unit carries the functional information, in this case with regard to tense, number and person.

(7)

நான்	காரை	வாங்கினேன்
naan	car-ai	vanginen
I.NOM	car-ACC	buy.PAST.1S

'I bought the car.'

(8)

நான்	அவனுக்குக்	காரை	வாங்கிக்கொடுத்தேன்
naan	avanukku-k	car-ai	vangikkoduththen
I.NOM	he.DAT-SAN	car-ACC	buy.VP-SAN.give.PAST.1SG

'I bought him a car.'

The f-structure in (9) shows an analysis of the complex predicate in (8) and illustrates the monoclausal nature of the co-predication. The f-structural analysis follows conventions established by the Urdu ParGram grammar (Butt & King 2007) and includes information about lexical semantics, which in this case only involves the information that the overall predicate is agentive. The co-predication is shown via the composed verbal PRED value, which brings together information contributed by each of the predicates.

(9)

$$
\left[\begin{array}{ll}
\text{PRED} & \text{`}\textsc{kodu}\left\langle(\uparrow \text{OBJ-TH}),\textsc{vangu}\left\langle(\uparrow \text{SUBJ}),\ (\uparrow \text{OBJ})\right\rangle\right\rangle\text{'} \\
\text{SUBJ} & \left[\begin{array}{ll}
\text{PRED} & \text{`}\textsc{pro}\text{'} \\
\text{PRON-FORM} & \textsc{naan} \\
\text{CASE} & \textsc{nom} \\
\text{NUM} & \textsc{sg} \\
\text{PERS} & 1
\end{array}\right] \\
\text{OBJ-TH} & \left[\begin{array}{ll}
\text{PRED} & \text{`}\textsc{pro}\text{'} \\
\text{PRON-FORM} & \textsc{avan} \\
\text{CASE} & \textsc{dat}
\end{array}\right] \\
\text{OBJ} & \left[\begin{array}{ll}
\text{PRED} & \text{`}\textsc{car}\text{'} \\
\text{CASE} & \textsc{acc} \\
\text{DEF} & +
\end{array}\right] \\
\text{TNS-ASP} & \left[\begin{array}{ll}
\text{TENSE} & \textsc{past} \\
\text{MOOD} & \textsc{indicative}
\end{array}\right] \\
\text{LEX-SEM} & \left[\begin{array}{ll}
\text{AGENTIVE} & +
\end{array}\right] \\
\text{VTYPE} & \left[\begin{array}{ll}
\text{COMPLEX-PRED} & \textsc{vv}
\end{array}\right] \\
\text{STMT-TYPE} & \textsc{decl} \\
\text{PASSIVE} & -
\end{array}\right]
$$

The example in (10) shows an alternative version of (8) in which the two parts of the complex predication are realised separately. The nature of the monoclausal co-predication at f-structure does not change with this alternative realisation. The *Sandhi* [k] is also triggered on the main verb வாங்கு (vangu) 'buy' just as in the single word realisation in (8). We discuss how this implementational challenge is resolved in section 4.2.

(10)

நான்	அவனுக்குக்	காரை	வாங்கிக்	கொடுத்தேன்
naan	avanukku-k	kar-ai	vangi-k	koduththen
I.NOM	he.DAT-SAN	car-ACC	buy.VP-SAN	give.PAST.1SG

'I bought car for him.'

3.3 Causatives

A causative in Tamil can be realised either morphologically or syntactically (Nuhman 1999, Paramasivam 2011). In either case, the causative is realised as a monoclausal complex predication (Annamalai 2013).

3.3.1 Morphological Realisation of Causatives

The causative can be realised in the morphology through three morphs: வி (vi), and (ப்)பி ((p)pi), which occur before the tense maker in a verb (Steever 2005). The choice of causative morph depends on the last vowel of the verb root and is thus phonologically conditioned.

The example in (11) shows how the morpheme *vi* is used in causatives. As shown in (12), the f-structure for (11) analyses the morphological causative as a CP (Butt 2010, Butt et al. 2003); the argument roles in this causative are causer.NOM and causee.INST; the case marking of the patient is NOM. The causative morpheme co-predicates together with the main verb.

(11)

வாங்குவித்தான் (vanguvittaan)

வாங்கு	-வி	-த்	-த்	-ஆன்
vangu	-vi	-t	-t	-aan
buy	-CAUS	-SAN	-PAST	-3SMR

'he made somebody buy (something)'

(12) $$\left[\text{PRED} \quad \text{`CAUS} \left\langle (\uparrow \text{SUBJ}), \text{`VANGU} \left\langle (\uparrow \text{OBL-INST}), (\uparrow \text{OBJ}) \right\rangle\text{'} \right\rangle\text{'}\right]$$

The example in (13) shows that LV constructions and causatives can be stacked in the sense that the causative applies to the entire LV construction, irrespective of whether the two verbs are written together or separately. The corresponding f-structural analysis is shown in (14).

(13)

நான்	அவளைக்கொண்டு	அவனுக்கு	
naan	avalaikkodu	avanukku	
I.NOM	she.ACC.INST	he.DAT	
ஒரு	கார்	வாங்கிக்	கொடுப்பித்தேன்
oru	car	vangi-k	koduppitthen
a	car.NOM	buy.VP-SAN	give.CAUS.PAST.1SG

‘I got her to buy a car for him.’

(14)

$$
\left[\begin{array}{ll}
\text{PRED} & \text{‘CAUS}\left\langle(\uparrow \text{SUBJ}),\text{‘KODU}\left\langle(\uparrow \text{OBJ-TH}),\text{‘VANGU}\left\langle(\uparrow \text{OBL-INST}),(\uparrow \text{OBJ})\right\rangle\text{’}\right\rangle\text{’}\right\rangle\text{’} \\
\text{SUBJ} & \left[\begin{array}{ll}\text{PRED} & \text{‘PRO’} \\ \text{PRON-FORM} & \text{NAAN} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 1\end{array}\right] \\
\text{OBJ-TH} & \left[\begin{array}{ll}\text{PRED} & \text{‘PRO’} \\ \text{PRON-FORM} & \text{AVAN} \\ \text{CASE} & \text{DAT}\end{array}\right] \\
\text{OBL-INST} & \left[\begin{array}{ll}\text{PRED} & \text{‘PRO’} \\ \text{PRON-FORM} & \text{AVAL} \\ \text{CASE} & \text{INST}\end{array}\right] \\
\text{OBJ} & \left[\begin{array}{ll}\text{PRED} & \text{‘CAR’} \\ \text{CASE} & \text{NOM} \\ \text{DEF} & -\end{array}\right] \\
\text{TNS-ASP} & \left[\begin{array}{ll}\text{TENSE} & \text{PAST} \\ \text{MOOD} & \text{INDICATIVE}\end{array}\right] \\
\text{VTYPE} & \left[\begin{array}{ll}\text{COMPLEX-PRED} & \text{VV}\end{array}\right] \\
\text{PASSIVE} & - \\
\text{STMT-TYPE} & \text{DECL}
\end{array}\right]
$$

3.3.2 Syntactic Realisation of Causatives

Causatives in Tamil can also be realised syntactically by adding one of the following verbs after an infinitive form of the main verb: செய் (sei) ‘do’, வை (vai) ‘put’, பண்ணு (pannu) ‘do’. These verbs do not predicate as full verbs in this case and have the character of light verbs, expressing the non-referential meaning ‘make’. When one of these three verbs is used as the main verb of a predication, the other two verbs can function as causativising light verbs in the combination. As shown in (15), we again face the implementational challenge that the main verb and the causative light verb can be written together as one token or separately as two tokens. Further, ப் (p) or ச் (c) will be added as a *Sandhi* to the main verb as part of the causativisation, for *pannu* 'do' and *sei* 'do', respectively. There is no *Sandhi* for *vai* 'put' as it begins with a consonant வ்(v).

(15)

அவனை	ஒரு	கார்	வாங்கச்	செய்தேன்
avan-ai	oru	car	vanga-c	seithen
he-ACC	a	car.NOM	buy.INF-SAN	make.PAST.3SM

'(I) made him buy a car.'

3.4 Passives

Passive constructions in Tamil are also realised via a V-V construction. The verb படு (padu) 'be touched/be experienced/sleep' is used in the passive constructions, where *padu* is an auxiliary verb. Together with an infinitive form of a main word, it gives the meaning of 'be subjected to' (Annamalai 2013).

For instance, consider (16) and its passive version in (17). As per the standard LFG lexical rule for passivisation, the original OBJ becomes the nominative SUBJ and the original SUBJ is realised as an instrumental adjunct (OBL-INST). The monoclausal analysis of the passive is shown below. As in causatives, CP passives can also be written as one word, for our example in (17) this is: வாங்கிக்கொடுக்கப்பட்டது (vangkikkodukkappaddathu). Passives can also be written as separate words as shown in (18). However, when 'give' is a light verb, according to the corpus analysis, the passive part is always written together with it as in (17).

(16)

ராம்	அவனுக்கு	ஒரு	கார்	வாங்கிக்	கொடுத்தான்
ram	avanukku	oru	car	vangki-k	koduththaan
ram.NOM	he.DAT	a	car.NOM	buy.VP-SAN	give.VP.PAST.3SG

'Ram bought a car for him.'

(17)

ராமால்	அவனுக்கு	ஒரு	கார்	வாங்கிக்	கொடுக்கப்பட்டது
ramaal	avanukku	oru	car	vangki-k	kodukkappaddathu
ram.INST	he.DAT	a	car.NOM	buy.VP-SAN	give.INF.SAN.PASS

'A car was bought for him by Ram.'

(18)

கொடுக்கப்	பட்டது
kodukkap	paddatu
give.INF.SAN	do.PAST.3SG
‘was given’	

(19)

$$\begin{bmatrix} \text{PRED} & \text{‘KODU}\left\langle(\uparrow \text{OBJ-TH}),\text{VANGU}\left\langle(\uparrow \text{SUBJ})\right\rangle\right\rangle\text{’} \\ \text{OBJ-TH} & \begin{bmatrix} \text{PRED} & \text{‘PRO’} \\ \text{PRON-FORM} & \text{AVAN} \\ \text{CASE} & \text{DAT} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘CAR’} \\ \text{CASE} & \text{NOM} \\ \text{DEF} & - \end{bmatrix} \\ \text{OBL-INST} & \begin{bmatrix} \text{PRED} & \text{‘RAM’} \\ \text{CASE} & \text{INST} \end{bmatrix} \\ \text{TNS-ASP} & \begin{bmatrix} \text{TENSE} & \text{PAST} \\ \text{MOOD} & \text{INDICATIVE} \end{bmatrix} \\ \text{VTYPE} & \begin{bmatrix} \text{COMPLEX-PRED} & \text{VV} \end{bmatrix} \\ \text{STMT-TYPE} & \text{DECL} \\ \text{PASSIVE} & + \end{bmatrix}$$

4 Grammar Implementation

4.1 Existing ParGram Strategies

A combination of the Restriction Operator (Kaplan & Wedekind 1993) and the CHECK feature have been used to handle complex predicates within ParGram, especially for causation and passivisation in languages that work similarly, but not identically to Tamil: Turkish (Çetinoğlu 2009) and Urdu (Bögel et al. 2019). We propose to use the ParGram framework and strategies to handle the Tamil V-V complex predication described in this paper and show how it can be extended to cover the morphonological challenges posed by external *Sandhi* phenomena.

The CHECK feature was introduced within ParGram as a way to handle well-formedness checking. Information that is only relevant for ensuring morphosyntactic well-formedness, but is not relevant for down-stream semantic interpretation or further “higher” Natural Language Processing applications can be stored here.

The Restriction Operator allows for the manipulation of f-structural information. Atrribute-feature values may be "restricted" out as shown in the f-structures in (20) and (21) where the CASE feature has been restricted out of (20) via an application of the Restriction Operator '/': $(\uparrow/\textsc{case})$. This type of restricting out might become necessary if a value for CASE had already been assigned by one part of the grammar but needed to be changed by another part. This situation arises, for example, if a main verb with a certain lexical semantic specification enters into a complex predication and the lexical semantics of the main predication "change" under complex predicate composition.

(20) $$\begin{bmatrix} \text{PRED} & \text{`}\textsc{Ram}\text{'} \\ \text{PERS} & 3 \\ \text{NUM} & \textsc{sg} \\ \text{CASE} & \textsc{nom} \end{bmatrix}$$

(21) $$\begin{bmatrix} \text{PRED} & \text{`}\textsc{Ram}\text{'} \\ \text{PERS} & 3 \\ \text{NUM} & \textsc{sg} \end{bmatrix}$$

Most importantly, however, the Restriction Operator provides the ParGram grammars with a means of combining PRED information as described in section 3 for the V-V complex predicates and causativisation in the absence of an implementation of Mapping or Linking Theory.[4] The use of the Restriction Operator for complex predicate formation (Butt et al. 2003, Butt & King 2003, Butt et al. 2008) and passivisation (Wedekind & Ørsnes 2003) has been described elsewhere and is not repeated here.

In the next section we show how we used these feature and design recommendations and extended them where necessary to develop the Tamil grammar. We have developed our own FSM analyser and generator (Sarveswaran et al. 2018, 2019) and this is integrated into our grammar development efforts and extended where necessary to meet the implementational challenges.

[4]There are no technical impediments to a implementation of Mapping or Linking Theory. However, in the absence of a theoretical consensus on the issue, the Restriction Operator has emerged as a mathematically well-defined way of dealing with predicate composition without affecting the underlyingly monotonic nature of the LFG implementation (Kaplan & Wedekind 1993). XLE operates with a rudimentary version of argument structure internally that bears some resemblance to Kibort's recent linking proposals (Kibort 2013, 2014, Crouch et al. 2017). This topic provides a potential for further fruitful exploration, but is out of the scope of this current paper.

4.2 Implementation and challenges

In this section, we discuss how we handle the interaction between *Sandhi*, complex predicates, causatives and passives in our Tamil grammar. We treat light verbs via the Restriction Operator in combination with morphophonological rules implemented in the FSM analyser ThamizhiFST (Sarveswaran et al. 2018, 2019). The CHECK feature was originally defined within ParGram to ensure of morphosyntactic wellformedness. Our treatment extends this feature to include the relevant *Sandhi* information.

4.2.1 Handing *Sandhi*

As shown in the previous sections, Tamil has two types of *Sandhi*, internal and external. Internal *Sandhi* can be handled entirely within our Tamil FSM ThamizhiFST. The treatment of internal *Sandhi* is fairly straightforward in that words with incorrect *Sandhi* patterns are not analysed and thus identified as misspellings.

However, external *Sandhi* has to be dealt with carefully. The morphological analyser is able to show whether a given word has a *Sandhi* letter at the end or not. However, it cannot check whether the *Sandhi* was used appropriately since that information will only become available as part of the syntactic analysis. We therefore extended the CHECK feature to check on the wellformedness of the morphophonology by ensuring that the correct *Sandhi* letter is indeed used.[5]

As shown in (22) for words with a *Sandhi* க் (k), we associate the f-structural information (↑CHECK _Sandhi-k) = + with the morphological tag provided by the FSM (see Kaplan et al. (2004) for details on the integration of an FSM with XLE). The morphological tag is part of the analysis of the Tamil word. Instances of Sandhi-p, Sandhi-c, and Sandhi-t are treated similarly.

(22) +Sandhi-k (↑CHECK _Sandhi-k) = +

This f-structural attribute is used to constrain the possible syntactic analyses in the grammar and to check whether the correct *Sandhi* has indeed been used in the syntactic context across two separate syntactic tokens.

[5]One reviewer suggested that *Sandhi* can be treated along the lines of initial mutations in Welsh as proposed by Mittendorf & Sadler (2006), through just the use of the morphological analyser, avoiding the need of a CHECK feature in the syntax. However, the Welsh initial mutation system does not follow from synchronically regular phonological rules, unlike in Tamil. In Tamil we also derive important syntactic information from the application or non-application of external *Sandhi* (e.g., adjectives vs. relative clauses). We therefore have decided to continue handling external *Sandhi* via a combination of FSM and the CHECK feature in the syntax.

4.2.2 Handling Complex Predicates in Tamil

When a CP is written as one token as in வாங்கிக்கொடுத்தேன் (vankikko-dutteen) ‘I bought for someone’, we provide a subcategorisation frame as part of the lexical rule and handle it as a regular lexical item.

However, when a CP is written as two tokens as in (13), the two predicational units can be composed via the Restriction Operator. We follow the analyses established in Butt et al. (2003) and Butt & King (2006) whereby the light verb subcategorises for regular arguments as well as a variable %PRED2. This variable will be substituted in by the subcategorisation frame of the main verb as part of the complex predicate composition. For instance, in our grammar, we have a lexical entry for a light verb with its functional features கொடுத்தேன் (kodutteen) ‘I gave’, as shown in (23).

We then use the template in (24) which we obtained from Dalrymple et al. (2004) in the grammar rules for light verb to compose arguments.

(23)

கொடு Vlight XLE (↑ PRED)=‘கொடு<(↑OBJ-TH) %PRED2>’ .

(24)

VV-ANNOTATION =	
(↓ CHECK _RESTRICTED)	= +
(↑ PRED ARG2)	= (↓ PRED)
↓\PRED\SUBJ\CHECK\ OBJ-TH\OBJ\PASSIVE	= ↑\PRED\SUBJ\CHECK\ OBJ-TH\OBJ\PASSIVE
(↑ OBJ)	= (↓ OBJ)
(↑ SUBJ)	= (↓ SUBJ)
(↑ VTYPE COMPLEX-PRED)	= vv.

4.2.3 Handling Passives

Recall that the causative and passive are monoclausal constructions, but that the two predicational heads can be written either as one token or as separate tokens. If the two verbs are written together, they can be dealt with by the morphological analyser as shown in (25).[6] The surface form of the word is associated with tags shown in (25) via a series of rules. The analysis provides the information that this is a passive verb in the past and that it is combined with a ‘give’ light verb.

[6]Effects of assimilation and Sandhi are shown within parentheses.

(25)

வாங்கிக்கொடுக்கப்பட்டது
vangikkodukkappattathu
vang-i-(k)kodu-(k)k(ap)-pattatu
vangu +verb +vp +give +past +pass

The stem (vangu) and the tag for the light verb ‘give’ are straightforwardly associated with subcategorisation information via the stem lexicon contained in the grammar, as shown in (23) and (26).

(26) வாங்கு V XLE (↑ PRED)=‘வாங்கு<(↑ SUBJ)(↑ OBJ)>’ .

Most of the attendant morphological tags are also straightforwardly associated with the corresponding f-structure information, e.g., (↑TNS-ASP) = PAST for the +past tag (Kaplan et al. 2004). However, an interesting question arises with regard to the treatment of the passive.

Classically, the passive is treated via lexical rules Bresnan (1982) in the ParGram grammars. These lexical rules are coded as part of the lexical entries, allowing for either an active or a passive version of the verb. The passivised version of vangu ‘buy’ in (27), for example, would result in the subcategorisation frame (↑ PRED) = ‘vangu<(↑SUBJ)>’ due to the application of the passive lexical rule.

(27) vangu	(↑ PRED) = ‘vangu<(↑SUBJ) (↑OBJ)>’	
+past	(↑ TNS-ASP TENSE) = PAST	
+3sg	(↑PERS) = 3	
	(↑NUM) = sg	

As outlined in section 3, passivisation can also be applied to a composed complex subcategorisation frame. For instance, the composed subcategorisation frame of the complex predicate வாங்கிக்கொடுத்தான் (vankikkoduthaan) ‘(he) bought’ is ‘give-buy <(↑OBJth) (↑SUBJ) (↑OBJ)>’. When it is passivised, the resulting subcategorisation frame would be ‘give-buy <(↑OBJth) (↑SUBJ)>’. Passivisation via lexical rules is straightforwardly implementable when the parts of the verbal sequence are contained within one lexical item. In this case the lexical rule can be applied to the whole composed subcategorisation frame that is coming out of the lexicon.

However, an analysis via a passive lexical rule is not possible when the predicational heads are realised as two separate tokens. This is because the passive morpheme is morphologically attached to only one of the items in the verbal sequence and would naturally apply to only the subcategorisation frame of that item (Çetinoğlu 2009). However, passivisation actually needs to be applied to the composed subcategorisation frame of the complex predicate (which is distributed across two separate tokens in this case).

We therefore modified the existing lexical rule treatment of passivisation and instead developed an analysis in terms of the Restriction Operator, as shown in (28) (cf. also Wedekind & Ørsnes (2003)). The advantage of this analysis is that all operations or subcategorisation frames are now treated via the same mechanism and predicate composition can be treated in the same way irrespective of how the parts of a complex predication are realised: as two separate tokens or as a single complex verb, expressing the intuition that morphological and syntactic complex predication involves the same mechanism (Alsina & Joshi 1991).

5 Conclusion

This paper has examined the interaction between Tamil benefactive V-V complex predicates, causativisation, passivisation and attendant morphophonological *Sandhi* effects in the context of computational grammar development within LFG. Tamil orthography provides a particular challenge for grammar development as the verbal sequences can be optionally realised as one single token or several different tokens, but that the externaal *Sandhi* effects surface in either case. We showed how these external *Sandhi* effects can be dealt with via an interaction between the FSM and the grammar by utilising the concept of CHECK features introduced within ParGram and extending it to checking morphonological well-formedness. We further showed modeled the interaction of V-V benefactive complex predicates with causation via the Restriction Operator, but encountered problems when we attempted to handle the interaction with passivisation via classical lexical rules. We thus proposed to handle passivisation via the Restriction Operator as well, lending support to previous analyses along these lines for Urdu and Turkish (Çetinoğlu 2009, Butt et al. 2003) as well as Danish (Wedekind & Ørsnes 2003).

References

Agesthialingom, S. 1971. A Note on Tamil Verbs. *Anthropological Linguistics* 121–125.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar*. Stanford: CSLI Publications.

Alsina, Alex, Joan Bresnan & Peter Sells (eds.). 1997. *Complex predicates: Structure and theory*. Stanford: CSLI Publications.

Alsina, Alex & Smita Joshi. 1991. Parameters in causative constructions. In *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, 1–15.

Annamalai, E. 2013. The Variable Relation of Verbs in Sequence in Tamil.

NINJAL International Symposium 2013: Mysteries of Verb-Verb Complexes in Asian Languages.

Annamalai, E, A Dhamotharan & A Ramakrishnan. 2014. *Akarātiyiṉ putiya patippil taṟkālat tamiḻ ilakkaṇa viḷakkam* xxxi–xlvii. Crea-A Publishers, India.

Bögel, Tina, Miriam Butt & Tracy Holloway King. 2019. Urdu Morphology and Beyond: Why Grammars Should not live without Finite-State Methods. In Cleo Condoravdi & Tracy Holloway King (eds.), *Tokens of Meaning: Papers in Honor of Lauri Karttunen*, 439–465. Stanford: CSLI Publications.

Bresnan, Joan. 1982. The Passive in Lexical Theory. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 3–86. The MIT Press.

Butt, Miriam. 1994. Machine translation and complex predicates. In H. Trost (ed.), *Proceedings of the Conference KONVENS '94 "Verarbeitung natürlicher Sprache"*, 62–71. Springer, Berlin.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam. 2010. The Light Verb Jungle: Still Hacking Away. In M. Harvey M. Amberber & B. Baker (eds.), *Complex predicates in cross-linguistic perspective*, 48–78. Cambridge: Cambridge University Press.

Butt, Miriam & Tracy H. King. 2003. Grammar writing, testing and evaluation. In Ali Farghaly (ed.), *A Handbook for Language Engineers*, 129–179. Stanford: CSLI Publications.

Butt, Miriam, Tracy H. King & Gillian Ramchand. 2008. Complex predication: Who made the child pinch the elephant? In Linda Uyechi & Lian Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, Stanford: CSLI Publications.

Butt, Miriam & Tracy Holloway King. 2002. Urdu and the Parallel Grammar project. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization, COLING*, 39–45. Association for Computational Linguistics.

Butt, Miriam & Tracy Holloway King. 2006. Restriction for Morphological Valency Alternations: The Urdu Causative. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, 235–258. CSLI Publications.

Butt, Miriam & Tracy Holloway King. 2007. Urdu in a Parallel Grammar Development Environment. *Language Resources and Evaluation: Special Issue on Asian Language Processing: State of the Art Resources and Processing* 41.

Butt, Miriam, Tracy Holloway King & John T. Maxwell III. 2003. Complex Predicates via Restriction. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG03 Conference*, 92–104. Stanford: CSLI Publications.

Butt, Miriam, Tracy Holloway King, Maria-Eugenia Nino & Frederique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Butt, Miriam & Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135. 7–29.

Çetinoğlu, Özlem. 2009. *A large scale LFG grammar for Turkish*: Sabanci University dissertation.

Chakrabarti, Debasri, Vaijayanthi Sarma & Pushpak Bhattacharyya. 2007. Complex predicates in Indian Language Wordnets. *Lexical Resources and Evaluation Journal* 40(3-4).

Crouch, Richard, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2017. XLE documentation. Palo Alto Research Center. http://www.parc.com/istl/groups/nltt/xle/xle_toc.html.

Dalrymple, Mary, Ronald M Kaplan, Tracy Holloway King et al. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, 199–208. Stanford: CSLI Publications.

Fedson, Vijayarani Jotimuttu. 1981. *The Tamil serial or compound verb*: University of Chicago, Department of Linguistics dissertation.

Hart, George L. 2000. Statement on the Status of Tamil as a Classical Language. https://southasia.berkeley.edu/tamil-classes.

Hulden, Mans. 2009. FOMA: a finite-state compiler and library. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 29–32. Association for Computational Linguistics.

Kaplan, Ronald M., John T. Maxwell III, Tracy Holloway King & Richard Crouch. 2004. Integrating Finite-state Technology with Deep LFG Grammars. In *Proceedings of the European Summer School on Logic, Language and Information (ESSLLI) Workshop on Combining Shallow and Deep Processing for NLP*, .

Kaplan, Ronald M. & Jürgen Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings of the 6th Conference of the Association for Computational Linguistics European Chapter (EACL)*, 193–202. Utrecht University.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, Stanford: CSLI Publications.

Kibort, Anna. 2014. Mapping out a construction inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, 262–282. Stanford: CSLI Publications.

Lahiri, Aditi & Jennifer Fitzpatrick-Cole. 1999. Emphatic Clitics and Focus Intonation in Bengali. In R. Kager & W. Zonneveld (eds.), *Phrasal* *phonology*, Dordrecht: Foris Publications.

Lehmann, Thomas. 1993. *A grammar of modern Tamil.* Pondicherry Institute of Linguistics and Culture, India.

Mittendorf, Ingo & Louisa Sadler. 2006. A Treatment of Welsh Initial Mutation. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, 343–364. Stanford: CSLI Publications.

Mohanan, Tara. 1994. *Argument Structure in Hindi.* Stanford: CSLI Publications.

Mohanan, Tara. 1997. *Multidimensionality of representation: NV complex predicates in Hindi.* Stanford: CSLI Publications.

Muthuchchanmugan. 2005. *Ikkala Mozhiyiyal (Modern Linguistics).* Mahin Press, Chennai, India.

Nuhman, M.A. 1999. *Basic Tamil Grammar (In Tamil).* Sri Lanka: Readers' Association.

Paramasivam, K. 2011. *Contemporary Tamil Grammar.* India: Adaiyaalam.

Pope, George Uglow. 1979. *A handbook of the Tamil language.* Asian Educational Services.

Sarveswaran, K, Gihan Dias & Miriam Butt. 2018. ThamizhiFST: A Morphological Analyser and Generator for Tamil Verbs. In *2018 3rd International Conference on Information Technology Research (ICITR)*, 1–6. IEEE.

Sarveswaran, K, Gihan Dias & Miriam Butt. 2019. Using meta-morph rules to develop morphological analysers: A case study concerning Tamil. In *Proceedings of the 14th International Conference on Finite-State Methods and Natural Language Processing*, 76–86. Dresden, Germany: Association for Computational Linguistics.

Senavaraiyar. 1938. *Tholkappiyam - Eluththathikaram.* Chunnakam, Sri Lanka: Thirumahal Press.

Steever, Sanford B. 1987. *The Serial Verb Formation in the Dravidian Languages.* Delhi: Motilal Banarsidass.

Steever, Sanford B. 2005. *The Tamil Auxiliary System.* New York: Routledge.

Thesikar, Sangara Namashivaya. 1957. *Nannool Viruthiyurai.* Chennai, India: Vithiyanubalana Press.

Wedekind, Jürgen & Bjarne Ørsnes. 2003. Restriction and Verbal Complexes in LFG: A Case Study for Danish. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG03 Conference*, 424–450. Stanford: CSLI Publications.

Left Dislocation in Hungarian

Péter Szűcs
University of Debrecen

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 293–313

http://csli-publications.stanford.edu/LFG/2019

Keywords: left dislocation, left peripheral constructions, syntactic integration, Hungarian

Szűcs, Péter. 2019. Left Dislocation in Hungarian. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 293–313. Stanford, CA: CSLI Publications.

Abstract

This paper investigates left dislocation constructions in Hungarian (whereby some discourse-prominent entity is placed at the left periphery of the clause, with a subsequent co-referential pronoun). Two subtypes are distinguished: a) "topic left dislocation", which is a syntactically integrated construction, sharing properties with Germanic-type left dislocation and clitic left dislocation in certain Romance languages; b) "free left dislocation", which is a loosely integrated structure, similar to various hanging topic left dislocations. The paper explores the structures' morphosyntactic and semantic properties, how an LFG-theoretic account of them can be formulated as well as the cross-linguistic implications of Hungarian left dislocation.

1. Introduction

Left dislocation (LD)[1] is a common label for constructions whereby some discourse-prominent entity is placed at the left periphery of the clause, with a subsequent co-referential pronoun. The term itself originates in Ross (1967), who used it for sentences like (1). As usual in the literature, "left dislocation" will be used as a descriptive label here, without commitment to a particular analysis. Furthermore, I will use the label "host" for the prominent entity itself (*John* in (1)) and "associated pronoun" or "pronominal associate" for the co-referential pronoun.

(1) *John$_i$, I like him$_i$.*

Since Ross's original analysis, a large body of literature has emerged about LD. Some of the most notable instances are Cinque (1977), the edited volume of Anagnostopoulou et al. (1997) and Grohmann (2003). There seems to be a consensus that at least two subtypes of LD should be distinguished. In one type of LD, there is some syntactic dependency between the host and the associated pronoun and the construction itself is properly (syntactically) integrated into the containing sentence. This LD is commonly referred to as "i-type" left dislocation. The second type of LD is thought of as a looser kind of dependency. There, the host and the pronoun are only related pragmatically, and the host itself is also assumed to be in some sense less integrated into the core clause structure. This LD is usually called "n-type" left dislocation (for "non-integrated").[2] Shaer (2009: 366) (2004) illustrates the two LD-types with

[1] I thank the participants of the LFG2019 Conference in Canberra for helpful suggestions. I am especially indebted to Louisa Sadler, Rachel Nordlinger and Ron Kaplan. I also thank my reviewers for their helpful comments. Naturally, any error is my responsibility.

The project no. 111918 (New approaches in the description of the grammar of Hungarian pronominals) has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K funding scheme.

[2] López (2016) refers to this theoretical distinction as "d-type" and "h-type" dislocations.

the following German sentence pair. Similar constructions have also been reported in other Germanic languages, e.g. Dutch, Icelandic.

(2)	a.	*Den*	*Hans*$_i$*,*	*den*$_i$	*jeder*	*mag*	.
		the.ACC	Hans	d-pron.ACC	everyone	likes	
	b.	*Der/*	*Den*	*Hans*$_i$*,*	*jeder*	*mag*	*ihn*$_i$*.*
		the.NOM	the.ACC	Hans,	everyone	likes	him
		'Hans, everyone likes him.'					(German)

(2a) is an i-type dislocation. The pronominal associate is a so-called weak d-pronoun, a kind of a demonstrative, which obligatorily matches the case of the host. In the n-type example in (2b), the associate is a personal pronoun. In addition to the obvious categorial and positional difference in (2b), there are other syntactic differences, for instance in (2b), case-matching is not obligatory. For the details of the German construction, the reader is referred to Frey (2004), the overall picture is that (2a) displays more "connectivity effects" than (2b). I will discuss related Hungarian data in the subsequent chapters. The English example in (1) (sometimes called "hanging topic left dislocation" (HTLD)) is usually analyzed as an n-type LD.

A related construction is clitic left dislocation (CLLD), which is standardly analyzed as falling into the i-type LD category. Its most obvious feature is that the pronominal associate is not a full personal or a demonstrative pronoun, but a weak form, a clitic. The Greek example in (3) is from Alexiadou (2006). CLLD has also been reported in Italian, Spanish and other Romance languages.

(3)	*Ton*	*Jani*$_i$	*den*	*ton*$_i$	*ksero.*
	the.ACC	John.ACC	NEG	clitic.ACC	know.1SG
	'John, I do not know him.'				(Greek)

The aim of the current paper is to investigate left dislocation constructions in Hungarian and to provide LFG-theoretic analyses for them. Besides, I will put Hungarian LD into a typological perspective.

The main claims of the paper are as follows:

i) Hungarian possesses both i-type and n-type left dislocations.
ii) The i-type left dislocation in Hungarian shows properties of both Germanic LD and CLLD.
iii) The n-type left dislocation in Hungarian is best analyzed as a "syntactic orphan", in the sense of Haegeman (1991) and Shaer (2009).
iv) Analyses consistent with the framework of LFG can be formulated about both types of Hungarian LD.

2. Left dislocation in Hungarian

Left-dislocation in Hungarian has been the subject matter of a number of papers. The most notable references are Kenesei et al. (1998), Lipták & Vicente (2009), Lipták (2010, 2012), Baloghné Nagy (2013) and den Dikken & Surányi (2017). In my discussion, I will build on these sources in terms of empirical background. However, as none of these are LFG-papers, my theoretical perspective will be different. As for the phrase structure of Hungarian, I align myself with the account of Laczkó (2017), where the sentence is headed by an iterative S-node, dominating a "topic-field" and a subsequent "quantifier field". Below these is the VP. The specifier of the VP may host some verbal modifier (preverb (PV in the glosses), negation, etc.) or a focussed constituent. The postverbal field has a non-configurational, flat c-structure.

(4) Hungarian clause-structure

a. [$_{S}$ *János*$_{TOPIC}$ [$_{S}$ *mindig*$_{QUANT}$ [$_{VP}$ *meg* [$_{V'}$ *ette* *az ebédet*.]]]]
John always PV ate.3SG the lunch.ACC
'John always ate the lunch.'

b. [$_{S}$ *János* $_{TOP}$ [$_{S}$ mindig $_{QUANT}$ [$_{VP}$ az ebédet$_{FOC}$ [$_{V'}$ ette meg.]]]]
John always the lunch ate.3SG PV
'It was the lunch that John always ate.'

In the following section, I will show that like other languages, Hungarian has two distinct LD-constructions. The i-type construction will be labelled "topic left dislocation" (TLD) as it is associated with (contrastive) topics and the n-type is going to be called "free left dislocation" (FLD). The latter is more flexible in terms of its syntax and information structure.

2.1. Topic left dislocation

2.1.1. Properties of TLD

The following sentence exemplifies topic left dislocation.

(5) *(Szerintem) Jánost*$_{i}$*, azt*$_{i}$ *meghívtuk.*
in.my opinion John.ACC that.ACC invited.1PL
'(I think) John, we invited him.'

As can be seen from the example above, there is a discourse-prominent entity (*Jánost* 'John.ACC'), which is followed by a demonstrative pronoun (*azt* 'that.ACC'). The pair is located in the topic-field of the sentence. As the left-peripheral adverb attests, the host does not have to be absolutely string-initial, as long as it is in the topic-field. Accordingly, quantified expressions are excluded from the construction. (Note that semantic considerations would also bar such constellations, see the discussion below about the referential properties of the pronoun).

(6) **[$_{QUANT}$ Sok ember] az hazament.*
many person that home.went.3SG

The host element and the pronoun are usually adjacent to each other but this is not a syntactic requirement, as (7) shows.

(7) *Jánost*$_{i}$*, Mari azt*$_{i}$ *meghívta.*
John.ACC Mary that.ACC invited.3SG
'John, Mary invited him.'

The construction is commonly associated with the contrastive topic discourse function. However, both Lipták (2012) and Baloghné Nagy (2013) mention that there exist sentences in Hungarian with LD which are not interpreted contrastively. Following them, I also do not consider TLD as necessarily contrastive, so neutral topics may also be involved. An example for this is (8), from Lipták (2012: 289). It has to be noted that in absence of knowing the communicative context and the speaker's intention, it is often hard to evaluate the contrastivity of a given utterance. Nevertheless, (8) does not feel contrastive at all. This is probably facilitated by the colloquial phrase "took himself and…", which gives the impression of a simple sequence of events.[3]

(8) *Erre Péter$_i$ az$_i$ fogta magát és elszaladt.*
then Peter that took.3SG himself and away.ran.3SG
'Then Peter, he went and ran away.'

Various lexical classes and grammatical functions may be included in TLD. (9a) illustrates this with an oblique complement, (9b) with an infinitive and (9c) with a predicative adjective.

(9) a. *A házban$_i$, [abban$_i$/ ott$_i$] nincs senki.*
the house.to that.in there not.be nobody
'The house, nobody is there.'

b. *Enni$_i$, azt$_i$ szeretek.*
eat.INF that.ACC like.1SG
'To eat, I like doing that.'

c. *Gazdag$_i$, az$_i$ nem vagyok.*
rich that not am
'Rich, I am not that.'

(9a) also shows that sometimes there is a choice with regards the demonstrative in TLD. The case-marked form of the basic demonstrative *az* 'that' is the standard option but if there is semantically matching specialized pronoun like the locative *oda* 'there' in the lexical inventory of the language, that may also be used. Thus *onnan* 'from.there', *oda* '(to) there', etc. are also available in the appropriate contexts.

It can be said that the choice of the demonstrative basically follows the pattern of general pronoun selection of Hungarian: whatever demonstratives would be selected in non-LD contexts, such pronouns are also utilized in Hungarian TLD.

However, there are some peculiarities. As shown in (5) above, personal names may be associated with a demonstrative pronoun in Hungarian TLD. However, in non-TLD contexts, such a reference would be considered

[3] A reviewer doubts (8) being non-contrastive. I disagree, though a lot depends on how one defines contrast. I think (8) includes a shifted topic, a new (or newly returned to) discourse referent which is different from clear cases of strong contrast, where there is an evoked set of contextually salient alternatives. For discussion of the notion of "contrast", see Repp (2016).

infelicitous, or at least impolite (regarding *John* not as a person but a thing), and a personal pronoun would be the default choice.

(10) Q: *Jánost hívtad meg?*
John.ACC invited.2SG PV
'Is it John that you invited?'
A: *Igen, [#azt/ őt].*
yes that.ACC him
'Yes, #that/ him.

However, this pragmatic infelicity is not felt in example (5), which indicates that the semantics/pragmatics of this LD-demonstrative is not completely identical to regular demonstratives.

Another point of divergence between regular demonstratives and the ones used in the TLD-construction has to do with number agreement. Interestingly, a plural host may be also associated with a singular TLD-pronoun. Such a pattern would not be possible in regular discourse using demonstrative pronouns.[4]

(11) a. *A fiúkat*$_i$*, [azt*$_i$ */ azokat*$_i$*] meghívtuk.*
the boys.ACC that.ACC those.ACC invited.1PL
'The boys, we invited them.'

b. Q: *A fiúkat hívtad meg?*
the boys.ACC invited.2SG PV
'Is it the boys that you invited?'
A: I*gen, [#azt/ azokat].*
yes that.ACC those.ACC
Approx.: 'Yes, I invited #him/them.'

(12) a. *A házakban*$_i$*, [abban*$_i$ */ azokban*$_i$*] nincs senki.*
the houses.in that.in those.in not.be nobody
'The houses, nobody is in them.'

b. Q: *A házakban nincs senki?*
the houses not.be nobody
'Is it the houses where there aren't anyone?'
A: *Igen, [#abban/ azokban].*
yes that.in those.in
'Yes, in #that/ those.'

The third interesting divergence from the standard usage of demonstrative pronouns is that a seemingly accusative-marked TLD-pronoun may be associated with a host that does not bear the OBJ grammatical function, as

---

[4] Notably, as Tibor Laczkó pointed out to me (p.c.), this pattern also surfaces with relative pronouns, especially in spoken language.

(i) *a fiúkat, akit / akiket meghívtam*
the boys.ACC whom.SG whom.PL invited.1SG
'the boys whom I invited'

shown in (13), from Lipták & Vicente (2009: 661). (13a) is the LD-structure and (13b) shows that the infinitival phrase must be the subject (and not the object) of the main predicate *jó* 'good'.

(13) a. *Úszni*$_i$, *az/ azt*$_i$ *jó volt*.
swim.INF that that.ACC good was
'To swim, that was good to do.'

b. *Jó volt [úszni/ az úszás/ *az úszást]*.
good was swim.INF the swimming.NOM the swimming.ACC
'Swimming was good.'

Apart from such special cases, the host and the pronominal associate show case-matching. (14) is the minimally modified version of (5). This is an instance of syntactic connectivity, noted in section 1.

(14) **János*$_i$, *azt*$_i$ *meghívtuk.*
John.NOM that.ACC invited.1PL
Intended: 'John, we invited him.'

Another instance of syntactic connectivity is variable binding. (15) shows that that a dislocate has no problem with being bound by a quantifier in the host clause.

(15) *A kutyá-já-t, azt mindenki szereti.*
the dog-POSS.3SG-ACC that.ACC everyone likes
'His (one's), dog, everyone likes that.'

The pronominal associate shows distal deixis by default. Proximal deixis is only possible if the host explicitly contains a proximal element. This is not a unique property of Hungarian TLD, the pattern shows up in other parts of Hungarian too, e.g. the pronominal associate of subordinate clauses is also distal by default.[5]

(16) **Jánost*$_i$, *ezt*$_i$ *meghívtuk.*
John.ACC this.ACC invited.1PL
Intended, approx.: 'John, we invited this one.'

(17) [*Ezt a fiút]*$_i$, *ezt*$_i$ *meghívtuk.*
this.ACC the boy this.ACC invited.1PL
'This boy, we invited him.'

(18) *Azt/ #ezt mondtam, hogy Jánost meghívtuk.*
that.ACC this.ACC said.1SG COMP John.ACC invited.1PL
'I said that we had invited John.'

After surveying the formal properties of the pronominal, let us now take a semantic perspective. From this angle, it is a crucial question point to settle whether the associated pronoun has a PRED feature or not. That is, should it be analyzed as having some sort of a reference or it is just a grammatical formative (expletive). I argue that the answer is the former, so the pronoun has semantic load and thus, a PRED feature. The arguments are as follows.

[5] For an overview of this construction, see Szűcs (2015).

Firstly, note the possibility of pronouns that are obviously semantically contentful, discussed in relation to (9a), *ott* ‘there’, *oda* ‘to.there’, etc. The spatial reference of these is quite recognizable, which fits much better with an approach where the LD-pronoun is not devoid of semantics.

Secondly, the TLD-pronoun induces semantic/pragmatic effects which are discernible in certain contexts. Basically it requires the host to be referentially anchored. Consider the examples in (19).

(19) a. *Valaki*$_i$ *(az*$_i$*)* *eljött.*
somebody that came.3SG
‘Somebody came.’

b. *Valaki*$_i$ *(*az*$_i$*)* *van odakint.*
somebody that is outside
‘There is somebody outside.’

(19a) can be interpreted if *valaki* ‘somebody’ refers to some contextually available set of people. For example, such a sentence may be used in a context like “We invited many people. Some of them came, some didn’t.”. (19b) is a presentational sentence, where the reference of *valaki* ‘somebody’ is newly introduced, so this anchored interpretation is not available. Accordingly, the use of the LD-pronoun is barred. Without it, (19b) is grammatical.

A similar contrast may be construed with *bárki* ‘anyone’. (20a) may be interpreted in a way that *bárki* ‘anyone’ is restricted to a certain group of people. (20b), where this anchored interpretation is not available, as the meaning unrestrictedly refers to people in general, is infelicitous.

(20) a. *Bárki*$_i$ *(az*$_i$*)* *nem jöhet be.*
anyone that not come.POT.3SG in
Intended: ‘Not just anyone may come in.’ (Lit.: ‘Anyone, they may not come in.’)

b. *Ha bárki*$_i$ *(*az*$_i$*)* *bejött, adtunk neki enni.*
if anyone that in.came.3SG gave.1PL him.DAT eat.INF
Intended: ‘If anyone came in, we gave them food.’ (Lit.: ‘If anyone, they came in, we gave them food.’)

Another indication of the semantic nature of this pronoun is its incompatibility with idiom-chunks. Consider (21).

(21) *A fene*$_i$ *(#az*$_i$*)* *megette ezt az egész ügyet.*
the heck that ate.3SG this the whole issue.ACC
‘This whole issue is screwed.’ (Lit.: ‘The heck, that ate this whole issue.’)

(21) is an intriguing sentence, as there is an idiom chunk in the topic field, which in itself should make the sentence anomalous, in theory. (Compare: *#The beans, John spilled (them).*) For some reason which is not really clear to me at this point, the pronoun-less version of the sentence is acceptable, even on the idiomatic reading. Several examples of this sort may be found via internet search. Whatever the reason for this is, adding the TLD-pronoun

makes the sentence semantically anomalous by forcing a degree referentiality on the subject phrase *a fene* 'the heck', which it is not compatible with.
It has to be noted that the force of this argument is somewhat diminished by the fact that splitting the idiom up by any means reduces the grammaticality of the sentence.

(22)	*A*	*fene$_i$*	*(?már)*	*megette*	*ezt*	*az*	*egész*	*ügyet.*
	the	heck	already	ate.3SG	this	the	whole	issue.ACC

'The whole issue was already screwed.' (Lit.: 'The heck already ate this issue.)'

However, while (22) with the interjecting *már* 'already' sounds marked, it is still not totally unacceptable, in contrast with the LD-version of (20). I take this as an indication that apart from the syntactic issue of breaking the continuity of an idiom, the semantics of the pronoun is also behind the problem in (20).

At this point it should be restated that TLD is not necessarily contrastive. This is important because otherwise one could argue that the explanation behind the data in (19)-(22) is simply the difficulty of construing contrastive readings for the sentences.

Additionally, I would like to call attention to Arregi (2003: 40), who describes similar effects in Spanish CLLD. In (23), *algo* 'something' may not be associated with the pronominal clitic *lo* 'it'.

(23)	*Algo$_i$,*	*Juan*	*si*	*(*lo$_i$)*	*comió.*
	something	Juan	yes	it	ate.3SG

'Something, Juan did eat.' (Spanish)

Arregi (2003: 40) argues that "the distribution of the clitic is determined by the interpretation of the clitic itself (…) In left dislocation, the clitic is interpreted as an individual variable". While the proper semantic/pragmatic characterization of the TLD-pronominal is yet to be worked out, it seems to be clear that it has to be interpreted some way, which precludes an analysis where it is an empty formative.

Another question about TLD is the nature of the relationship between the host and the pronoun: which of them is the dominant participant in the sentence? Here I agree with Zaenen (1997), who argues for an analysis of Icelandic left dislocation where the pronoun is an adjunct of the host. This is the most plausible analysis for Hungarian as well. The alternative is the reversed constellation, whereby the pronoun is the argument of the main predicate and the host is an adjunct, resembling an appositive construction. While such an analysis might be plausible for some Germanic TLD-constructions,[6] it is definitely not for Hungarian. To prove this, first recall the data from (13) where it is an infinitival complement that satisfies the

[6] Frey (2004) and Alexiadou (2006) propose analyses along this path. Whether Zaenen's (1997) analysis should be revised too is matter of further inquiry. I will explore some of the cross-linguistic and theoretical landscape of LD in section 3.

subcategorization requirement of the main predicate and not an accusative element like the LD-pronoun. Moreover, evidence for the primary status of the host over the pronoun can also be seen from object definiteness agreement patterns.

As illustrated in (24), finite verbs in Hungarian show definiteness agreement with their objects. Demonstrative pronouns count as definite objects, evidenced by (24).

(24) a. *Egy* *fiút* *lát-tál.*
one boy.ACC see-PAST.2SG.INDEF
'You saw a boy.'

b. *A* *fiút* *lát-tad.*
the boy.ACC see-PAST.2SG.DEF
'You saw the boy.'

(25) *Azt* *lát-tad/* **lát-tál.*
that.ACC saw-2SG.DEF see-PAST.2SG.INDEF
'You saw that.'

In an LD-sentence, it is always the host and not the pronominal associate that determines the definiteness agreement of the verb. Hence in (26a) the verb shows indefinite agreement, triggered by *egy fiút* 'a boy.ACC', even though there is the demonstrative LD-pronoun in the sentence, which in principle could trigger definite conjugation. (See also den Dikken & Surányi 2017: 571-572).

(26) a. *Egy* *fiút$_i$* *azt$_i$* *lát-tál.*
one boy.ACC that see-PAST.2SG.INDEF
'A boy, you saw him.'

b. *A* *fiút$_i$* *azt$_i$* *lát-tad.*
the apple.ACC that see-PAST.2SG.DEF
'The boy, you saw him.'

2.1.2. An LFG-approach to TLD

For Icelandic LD, Zaenen (1997) proposes an analysis whereby the pronoun is regarded as an adjunct of this topical host, as shown in (27).

(27) S → XP XP V NP
(↑TOP)=↓ (↑TOP-ADJ)=↓ ↑=↓ (↑SUBJ)=↓

Based on the considerations outlined above, I propose an analysis in a similar spirit. This is shown in Figure 1 for topic left dislocation in Hungarian, exposed via annotated phrase structure.

The pronominal associate is located in the topic-field of the Hungarian sentence, and the annotation for it should be optionally available (for details of Hungarian clause-structure, see Laczkó 2017). It is associated with some topical element, which is understood as covering contrastive and neutral topics alike.

The first line of the annotation of the TLD-pronoun is about providing its host with a "local name" (see e.g. Dalrymple 2001: 146-148) This is a formal device that makes it possible to refer to a particular f-structure in subsequent constraints. Here it singles out one a grammatical function, which is then identified as the "host" of the TLD-pronoun. The second line constrains the host to be a topic. Following the spirit of Zaenen's (1997) analysis, the pronoun is regarded as an adjunct of this host, as the equation in the second line of the annotation specifies. The constraining equation in line four requires this element to be an LD-pronoun. As argued earlier, I take these to be referential and their semantics should have commonalities with standard demonstratives but the data in (10)-(13) suggests that they should be treated separately. Line five requires co-reference between the host and the pronominal associate. Finally, the last line is about the case-requirements of the construction. In the default scenario, the host and the TLD-pronoun have matching case features, as evidenced by (14). Alternatively, the pronominal associate may lack a case feature, which happens for example with *ott* 'there' in (9a), or in instances where the host is not case-marked (e.g. (9b) or (13a)).

Two notes are due with respect to this last point, i.e. case. The first is that Zaenen (1997: 133) argues that case-matching follows from general rules in Icelandic, as adjuncts in Icelandic typically "agree in case marking, gender and number with the constituent they are an adjunct to", as e.g. in (28). As (29) shows, there is no such constraint in Hungarian (the form of *egyedül* 'alone' does not vary depending on the subject), that is why the matching has to be stated separately.

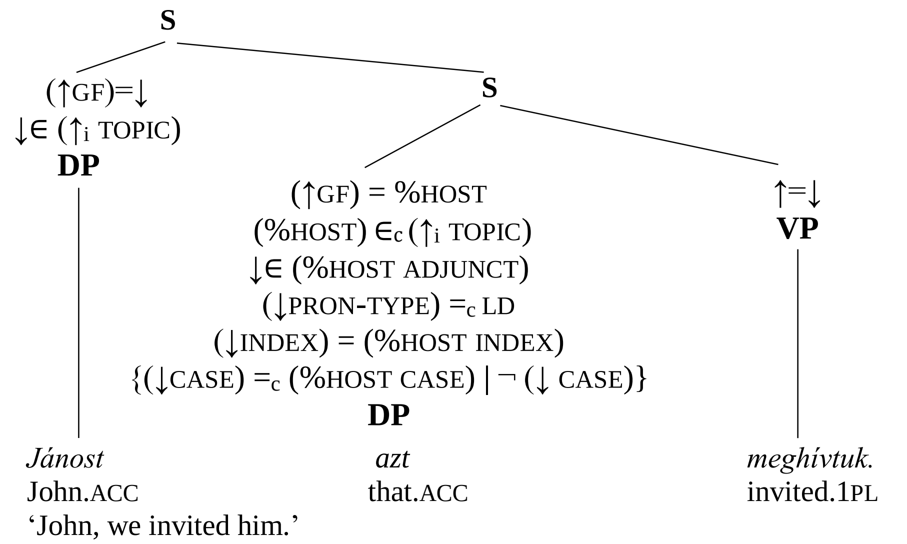


Figure 1.
TLD in Hungarian

(28)	*Ég*	*geri*	*petta*	*einn.*	
	I MASC.SG.NOM	will.do	this	alone.MASC.SG.NOM.	
	'I will do this alone.'				(Icelandic)

(29)	a.	*Én*	*ezt*	*egyedül*	*fogom*	*csinálni.*	
		I	this.ACC	alone	will.1SG	do.INF	
		'I will do this alone'					
	b.	*Ők*	*ezt*	*egyedül*	*fogják*	*csinálni.*	
		they	this.ACC	alone	will.3PL	do.INF	
		'They will do this alone.'					(Hungarian)

The second point is that I propose to handle case-discrepancies with alternate lexical entries for the respective pronouns. This differs from the approach of Lipták & Vicente (2009) and Lipták (2012), where predicate left dislocation (e.g. (13a)) is analyzed as being the result of a process that is distinct from other instances of TLD. Lipták & Vicente (2009) propose that the accusative case on the pronoun in (13a) is the manifestation of default case in Hungarian. In my approach, the accusative case is just apparent, this alternative lexical entry of the pronoun is caseless. I consider this to be a better approach as a unitary underlying mechanism is posited for all TLD-structures in Hungarian. Moreover, it is not evident that accusative is the default case in Hungarian, see e.g. (9c), where the adjective is associated with a nominative pronoun. It is also to be noted that in (13a), the nominative pronoun is still an equally valid option, which suggests that the accusative-marking may be misleading.[7]

Also, in contrast to English, left-peripheral, hanging pronouns are not in the accusative case, which argues against accusative being the default in Hungarian.[8]

(30) *Me, I like beer.*

(31)	*[Én/*	**Engem],*	*én*	*szeretem*	*a*	*sört.*
	I	me	I	like.1SG	the	beer.ACC
	'Me, I like beer.'					

This latter construction is distinct from TLD, it is an instance of free left dislocation, to which we turn in the next section.

---

[7] According to a reviewer, my approach is need of a stronger theoretical foundation. This may be true, but this is also true for the alternative, default case. Giving some formal substance to the theoretical notion of "default case", would have to resort to some mechanism that ensures that such an accusative case is not the same as standard accusative case. This is likely to result in something very close to what I propose.

[8] Bartos (2002, footnote 5) notes that the dative may surface in imperative root infinitives. This may also be regarded as some sort of a default case, crucially non-identical to the accusative.

(i)	*?A*	*fiúknak*	*leülni!*
	the	boys.DAT	sit.INF
	'Boys, sit!'		

A final point to make is that I assume that the LD-pronoun is specified for the person feature (3rd person), but the apparently singular one is underspecified with respect to number, which enables it to appear in sentences like (11) and (12).

2.2. Free left dislocation

2.2.1. Properties of FLD

(32) exemplifies what I label as free left dislocation (FLD).

(32) *Jánost*$_i$, *őt*$_i$ *meghívtuk.*
John.ACC him invited.1PL
'John, we invited him.'

In contrast to TLD, which prosodically forms a unit with the rest of the sentence, the left-peripheral element in FLD is set apart by a noticeable intonational break.

Another salient difference is that personal names are associated with personal pronouns, as one would expect in standard discourse. This feature of FLD can be most clearly explicated in conjunction with another property of the construction, the wider range of information structural categories that can be involved. In addition to the topic discourse function, the FLD pronoun can also be a focus of the main clause (first noted by Kenesei et al. 1998). This is seen in (33), where the focussed pronoun in the preverbal position pushes the preverb *meg* (contributing to the perfective interpretation of the sentence) to the postverbal field. In such cases, using a demonstrative like the ones in TLD triggers the sort of pragmatic infelicity demonstrated in (10) above.

(33) *Jánost*$_i$, *[*$_{VP}$ *őt*$_{i\ FOC}$ */* *#azt*$_{i\ FOC}$ *[*$_{V'}$ *hívtuk* *meg]].*
Johh.ACC him that.ACC invited.1PL PV
'John, we invited [HIM/#THAT].'

As for (the lack of) syntactic connectivity, consider (34), where the FLD example shows non-identical cases on the dislocate and the host. This contrasts with (14) above. (In 34a, the host is a topic, while in 34b, the host is a focus.)

(34) a. *János*$_i$, *őt*$_i$ *meghívtuk.*
John.NOM him invited.1PL
'John, we invited him.'

b. *János*$_i$, *őt*$_i$ *hívtuk* *meg.*
John.NOM him invited.1PL PV
'John, we invited him.'

The lack of connectivity may also be seen in example (35), contrasting with (15), where the binding of the (unexpressed) possessor by the quantifier is less than perfect.[9]

[9] That sentences like (35) are not entirely unacceptable could be a result of some poorly-understood processes that make variable binding possible even when the necessary syntactic configurations do not hold. In fact, such claims have also been

(35)	?*A*	*kutyá-já-t,*	*AZT*	*szereti*	*mindenki*
	the	dog-POSS.3SG-ACC	that.ACC	everyone	likes
	'His (one's), dog, everyone likes it.'				

Lastly, FLD contrasts with TLD in that it becomes marked if the host element is not string-initial. This obviously happens in subordinate clauses, but the same effect may appear in main clauses as well. Consider the FLD (a)- and TLD (b)-examples below.

(36)	a.	*?Mari*	*Jánosnak$_i$,*	*neki$_i$*	*adott*	*ajándékot.*
		Mary	John.DAT	him.DAT	gave.3SG	gift
	b.	*Mari*	*Jánosnak$_i$,*	*annak$_i$*	*adott*	*ajándékot.*
		Mary	John.DAT	that.DAT	gave.3SG	gift
		'John, Mary gave him a present.'				
(37)	a.	*?Mondtam,*	*hogy*	*Jánost$_i$,*	*őt$_i$*	*meghívtuk.*
		said.1SG	COMP	John	him	invited.1PL
	b.	*Mondtam,*	*hogy*	*Jánost$_i$,*	*azt$_i$*	*meghívtuk.*
		said.1SG	COMP	John	that.ACC	invited.1PL
		'I said that John, we invited him.'				

Also, (38) contrasts with (5), from section 2.1.1.

(38)	*?Szerintem*	*Jánost,*	*őt*	*meghívtuk.*
	in.my.opinion	John.ACC	him	invited.
	'I think John, we invited him.'			

2.2.2. An LFG-approach to FLD

Based on the considerations above, I argue that the most plausible analysis for FLD is one where the left-peripheral entity is syntactically independent from the rest of the sentence. In other words, it is regarded as a "syntactic orphan", using the terminology of Haegeman (1991) and Shaer (2009).[10] The relation between the host (the left-peripheral element) and the pronominal associate is like the relation between entities in two different utterances, a standard cross-sentential anaphoric dependency. This conception of FLD naturally explains the intonational break between the host and the sentence itself. Also, the use of personal pronouns in sentences like (28) is expected since they are the normal choice for such contexts. Given the pragmatic nature of the relationship, case-mismatches are also not a surprise.

Thus, from an LFG-perspective we need to find some mechanism allows a string to be analyzed as composed of independent substructures. For this, Fortmann's (2005) proposal about parenthetical expressions may be a path forward. What he proposes is that sequences like (39) should be analyzed in a

made in connection with English HTLD, see e.g. Vat (1981), who reports that (i) is not entirely ruled out.

(i) *?His$_i$ first article, I think [every linguist]$_i$ would consider it a failure.*

[10] For similar ideas, see for example Aissen (1992) and Banfield (1982). I thank one of my reviewers for these references.

way that the underlined segment is part of the c-structure of the entire expression, but it projects an independent f-structure.

(39) *Theo hat – der Klempner war nicht gekommen – die Heizung*
Theo has the plumber had not come the heating
repariert
fixed
'Theo has ((as) the plumber didn't come) the repaired the heating.'
(German)

The goal of projecting an independent f-structure is achieved by using the ↓=↓ notation for the parenthetical expression, instead of the standard ↑=↓ or (↑GF)=↓ equations. That is, the non-integrated element projects an f-structure, but this f-structure is not part of the f-structure of the host.

Thus, (32) should be analyzed as shown in Figure 2.

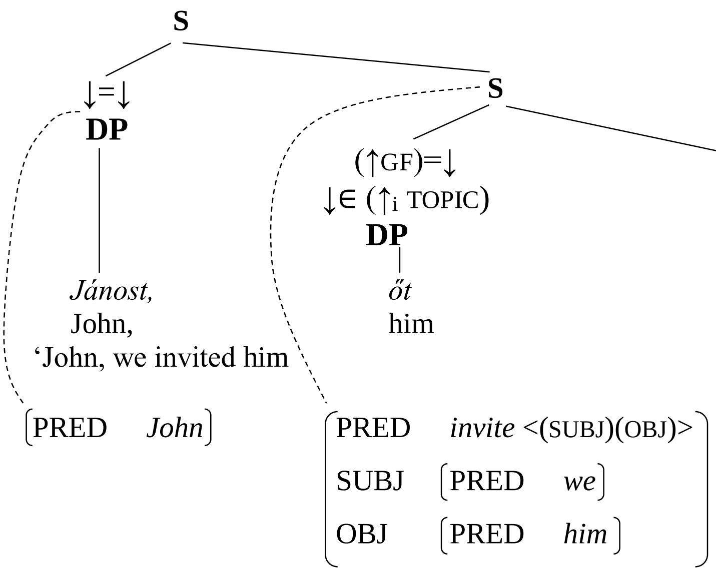


Figure 2.
FLD in Hungarian

The mild ungrammaticality of sentences like (36)-(38) then arguably comes from the extra-syntactic nature of the construction. This is possibly linked to processing factors, more precisely, from the difficulty of parsing phonologically intermingled independent utterances.

Finally, although such "hanging" left dislocation structures are usually associated with the topic discourse function, nothing in principle excludes other discourse functions to be associated with FLD. I will explore this and other typological aspects of left dislocation in some detail in the next section.

3. Typological considerations in left dislocation

In the previous section I gave an overview and possible LFG-theoretic approaches to left dislocation constructions in Hungarian, topic left dislocation (TLD) and free left dislocation (FLD). Now I turn to how these constructions compare to the typological landscape of LD, which was briefly outlined in the introduction.

As shown in (2), repeated here as (40), German also has two LD-constructions, which are commonly analyzed as i- and n-type left dislocations, respectively. Similar patterns have been described in Dutch and Icelandic, see the edited volume of Anagnostopoulou et al. (1997).

(40)	a.	*Den*	*Hans$_i$,*	*den$_i$*	*mag*	*jeder.*		(German)
		the.ACC	Hans	d-pron.ACC	likes	everyone		
	b.	Der/	*Den*	*Hans$_i$,*	*jeder*	*mag*	*ihn$_i$.*	
		the.NOM	the.ACC	Hans,	everyone	likes	him	
		‘Hans, everyone likes him.’						

From the discussion in the previous sections it is clear that Hungarian fits into this pattern, TLD being an i-type dislocation and FLD being an n-type one.

As such, TLD is given a syntactic analysis and it is properly integrated into the clause structure, as outlined in Figure 1. It utilizes demonstrative-like pronouns parallel to the the d-pronoun *den* in (40a), with syntactic restrictions on the formal features of this pronoun.

Semantic effects of the presence of the pronominal associate are also to be observed in German. Frey (2004: 214) exemplifies such effects with the following sentence.

(41)	Context: this is the children’s first day on their vacation.						
	Der	*Otto$_i$,*	*(der$_i$)*	*wollte*	*Fußball*	*spielen.*	(German)
	the.nom	Otto	d-pron	wanted	soccer	play	
	‘Otto, he wanted to play football.’						

Similarly to the observed effects in (19)-(20), if the LD-pronoun is present, *Otto* must be the member of some contextually given set of children. Without the pronoun, the referent may be newly introduced into the discourse.

As noted, my analysis for TLD is similar to that of the analysis of Zaenen (1997) for Icelandic LD. Frey (2004), in a Minimalist framework, also argues for an analysis of this sort, where the left-dislocated phrase and the pronoun are independently “base-generated” and co-indexed. However, in his account, the pronoun is in a theta-position and the left-peripheral phrase is a CP-adjunct, so the functional hierarchy is the opposite of Zaenen’s (1997) and mine. As already argued in section 2.1.1, while this might be the right approach for Germanic LD, it is definitely not the one for Hungarian. Apart from the arguments already mentioned, let us also note that the LD-pronoun by itself may be fully felicitous in German given the appropriate context (as in (42a)), this is not the case in Hungarian, as the demonstrative cannot refer to a person, except in the TLD construction, see (42b). As noted earlier, without the host

Jánost 'John.ACC', the pronoun could only refer to some nonhuman entity. Thus an analysis where the host is an adjunct is more plausible in Germanic LD than in Hungarian.

(42) a. *(Den Hans$_i$), den$_i$ mag jeder.* (German)
the.ACC Hans d-pron.ACC likes everyone
b. #*(Jánost$_i$), azt$_i$ mindenki kedveli.* (Hungarian)
John.ACC that.ACC everyone likes
'John, everyone likes him.'

The CP-adjoined position of the left-peripheral element in German is supported by the fact that it can marginally occur in a pre-complementizer position in a subordinate clause, as in (43a), from Frey (2004, footnote 14). This configuration is sharply ungrammatical in Hungarian, see (43b).

(43) a. *Maria glaubt, den Hans$_i$, dass den$_i$ jeder mag.*
Mary believes the.ACC Hans COMP d-pron everyone likes
'Maria believes that Hans, everyone likes him.' (German)
b. **Mari hiszi, Jánost$_i$ hogy azt$_i$ mindenki*
Mary believes John.ACC COMP that.ACC everyone
kedveli.
likes (Hungarian)

Another divergence from the Germanic pattern is that in these languages, LD is restricted to root clauses and subordinate clauses introduced by bridge verbs Frey 2004: 226). This is not the case in Hungarian, where TLD is freer in its distribution. This is evidenced by the contrast between the German and the Hungarian data below.

(44) a. **Maria bezweifelt, den Hans$_i$, dass den$_i$*
Mary doubts the.ACC Hans COMP d-pron
jeder mag.
everyone likes (German)
b. *Mari kétli, hogy Jánost$_i$, azt$_i$ mindenki szereti.*
Mary doubts COMP John.ACC that.ACC everyone likes
'Maria doubts that Hans, everyone likes him.' (Hungarian)

These data about subordinate clauses suggest that TLD in Hungarian is closer to the core sentential domain than the Germanic LD type. This likens the Hungarian construction to clitic left dislocation (CLLD) structures (see (3) above), which are analyzed as being IP-adjuncts by Alexiadou (2006). While on our approach, there is no IP in Hungarian, the parallel is that the construction is located in the standard sentential domain, which is IP in configurational languages and S in a language like Hungarian. This gives a straightforward explanation for the contrasts in (43)-(44).

It may be added here that since the topic field is inherently iterative in Hungarian, there is no point of talking about the host being an adjunct, in contrast to other instances of CLLD, noted above. This difference in phrase-structural configuration may be one of the reasons why the host is able to dominate the pronominal associate in terms of functional structure.

Another CLLD-like property of TLD is that it allows for stacking, which is not possible in the Germanic type of integrated LD. Consider the data in (45)-(47), where the non-Hungarian examples are from Alexiadou (2006). (41) is an Italian sentence which shows multiple instances of CLLD. (46) is Dutch LD demonstrating the ungrammaticality of multiple LDs. The Hungarian equivalent in (47) is possible.

(45)	*Di*	*vestiti*$_i$	*a*	*me*	*Gianni*$_j$	*in quel*	*negozio*$_k$	*non*	*mi*$_j$	
	DET	clothes	to me		Gianni	in that	shop	not	to.me	
	ce$_k$	*ne*$_i$	*ha*	*mai*	*comprati.*					
	there	of.them	has	ever	bought					

‘As for clothes, for me, Gianni has never bought them in that shop.’ (Italian)

(46)	**Jan*$_i$	*op school*$_j$	*die*$_i$	*daar*$_j$	*zag*	*ik*	*niet.*		
	John	at school	that	there	saw	I	not		(Dutch)

(47)	*Jánost*$_i$	*az iskolában*$_j$	*azt*$_i$	*ott*$_j$	*nem*	*láttam.*	
	John.ACC	the school.in	that.ACC	there	not	saw.1SG	

‘John, in the school, I didn’t see him there.’ (Hungarian)

Thus it seems that TLD is closer to CLLD constructions than Germanic LD, as far as syntactic distribution is concerned. However, the LD-pronoun in Hungarian is not a clitic, but a demonstrative-like element, like in the Germanic type.

FLD, just like the German example in (40b), involves a loosely attached left peripheral element which is only pragmatically related to the subsequent pronominal, which then may naturally be a personal pronoun. It was described in section 2.2.1 that FLD seems to be degraded in non-initial positions. Such a degradation may be observed with regards other loose attachment-constructions as well. English hanging topic left dislocation is a standard example for these. The picture is not uncontroversial (for different perspectives, see Grohmann 2003: 139 vs. Shaer 2009: 379), it is plausible to claim that the embedded HTLD in (44) deserves a question-mark. There is a related datum in (45), which points to the same direction. There, we see that HTLD may precede but not follow topicalization, the latter being a syntactically integrated long-distance dependency. Similar data is reported in German by Grohmann (2003: 148), shown in (50). (8) would also be fully grammatical as a topicalization structure (i.e. without the pronoun associate).

(48) *John said that Mary*$_i$*, he likes (?her*$_i$*).*

(49) a. *?Mary, John*$_i$*, he*$_i$ *likes.*

b. *Mary*$_i$*, John, she*$_i$ *likes.*

(50) a. **Einen Arschtritt dieser Kandidat*$_i$*, sollte man*
a.ACC kick-in-the-ass this.NOM candidate should one
ihm$_i$ *geben.*
him give
Intended: 'A kick in the ass, this candidate, one should give him.'
(German)

b. *Dieser Kandidat*$_i$ *einen Arschtritt, sollte man*
this.NOM candidate a.ACC kick-in-the-ass should one
ihm$_i$ *geben.*
him give
'This candidate, a kick in the ass, one should give him.' (German)

As noted earlier, n-type LDs are commonly associated with hanging topics. However, nothing conceptually excludes other discourse functions, so potential association with focus in Hungarian FLD just fills a typologically available but unattested scenario. TLD is tied to the topic discourse function, but again this is not a necessity for i-type dislocations. Both Grohmann (2003: 145) and Frey (2004: 213) assert that German left dislocation may be used as a contrastive focus. Thus, the inventory of information structural categories for LD constructions has to be established on the basis of individual languages.

(51) Q: *Have you met Anna yesterday?*

a. A: *Nein. Den Martin*$_i$*, den*$_i$ *habe ich gestern*
no the.ACC Martin d-pron.ACC have I yesterday
getroffen.
met (German)

b. A: *#Nem. Martinnal*$_i$*, azzal*$_i$ *találkoztam.*
no Martin.with that.with met.1SG
'No. I met Martin yesterday.' (Hungarian)

4. Conclusion

In this paper I gave an overview of left dislocation (LD) constructions in Hungarian, with a typological outlook. I argued that Hungarian follows the cross-linguistic pattern whereby LD bifurcates into a syntactically integrated (i-type) and a non-integrated (n-type) construction.

I labelled the i-type construction of Hungarian "topic left dislocation" (TLD), given its association with (contrastive) topics. It was given an LFG-analysis in the spirit of Zaenen (1997), whereby the pronominal associate is an adjunct of its host. The characteristics of the construction follow from the phrase-structural rules and the properties of the LD-pronoun itself. While the form of the pronominal likens TLD to Germanic left dislocation constructions, its syntactic distribution is more similar to clitic left dislocation.

The n-type construction, "free left dislocation" (FLD), is claimed to be a "syntactic orphan", an entity loosely attached to the sentence, akin to hanging topic left dislocation constructions. According to this view, the host and the pronominal are parts of the same c-structure, but project a separate f-structure, as Fortmann (2005) proposed for parenthetical expressions.

References

Aissen, Judith L. 1992. Topic and focus in Mayan. *Language* 68(1). 43-80.

Alexiadou, Artemis. Left dislocation (including CLLD). 2006. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. 2, 668-699. Blackwell Publishing.

Anagnostpoulou, Elena, Henk van Riemsdijk & Frans Zwarts (eds.). 1997. *Materials on left dislocation*. Amsterdam: John Benjamins.

Arregi, Karlos. 2003. Clitic left dislocation is contrastive topicalization. In *Proceedings of the 26th annual Penn Linguistics Colloquium, Penn working papers in linguistics* 9(1), 31-44. Philadelphia: Penn Linguistics Club.

Baloghné Nagy Gizella. *Left dislocation in Optimality Theory*. Doctoral dissertation, Eötvös Loránd University.

Banfield, Ann. 1982. *Unspeakable sentences*: *Narration and representation in the language of fiction*. Routledge & Kegan Paul.

Bartos Huba. 2002. Root infinitives. In Kenesei István & Siptár Péter (eds.), *Approaches to Hungarian 8*: *Papers from the Budapest conference*, 13-38. Budapest: Akadémiai Kiadó.

Cinque, Gugliemo. 1977. The movement nature of left dislocation. *Linguistic Inquiry* 8. 387-412.

den Dikken, Marcel & Surányi Balázs. 2017. Contrasting contrastive left dislocation explications. *Linguistic Inquiry* 48(4). 543-584.

Fortmann, Christian. 2005. On parentheticals (in German). In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG05 conference*, 166-185. Stanford, CA: CSLI Publications.

Frey, Werner. 2004. Notes on the syntax and pragmatics of German left dislocation. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*, 203-233. Berlin: Mouton de Gruyter.

Grohmann, Kleanthes K. 2003. *Prolific domains: on the anti-locality of movement dependencies*. Amsterdam: John Benjamins.

Haegeman, Liliane. 1991. Parenthetical adverbials: The radical orphan approach. In Shuji Chiba et al. (eds.), *Aspects of modern English linguistics: Papers presented to Masatomo Ukaji on his 60th birthday,* 232-254. Tokyo: Kaitakusha.

Kenesei István, Robert M. Vago & Fenyvesi Anna. 1998. *Hungarian*. London: Taylor & Francis Routledge.

Laczkó Tibor. 2017. *Aspects of Hungarian syntax from a lexicalist perspective*. Doctor of the Hungarian Academy of Sciences thesis. University of Debrecen.

Lipták Anikó & Luis Vicente. 2009. Pronominal doubling under topicalization. *Lingua* 119. 650–686.

Lipták Anikó. 2011. The structure of the topic field in Hungarian. In Beninca, Paola & Nicola Munaro (eds.), *Mapping the left periphery*, 163-198. Oxford: Oxford University Press.

Lipták Anikó. 2012. Correlative topicalization. *Acta Linguistica Hungarica* 59(3). 245-302.

López, Luis. 2016. Dislocations and information structure. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Repp, Sophie. 2016. Contrast: Dissecting an elusive information-structural notion and its role in grammar. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Ross, John R. 1967. *Constraints on variables in syntax*. Doctoral dissertation, Massachusetts Institute of Technology.

Shaer, Benjamin. 2009. German and English left-peripheral elements and the orphan analysis of non-integration. In Benjamin Shaer, Philippa Cook, Werner Frey & Claudia Maienborn (eds.), *Dislocated elements in discourse: Syntactic, semantic and pragmatic perspectives*, 366-397. New York: Routledge.

Szűcs Péter. 2015. On pronouns in Hungarian complex sentences. *Argumentum* 11. 292-313.

Vat, Jan. 1981. Left dislocation, connectedness, and reconstruction. *Groninger Arbeiten zur Germanistischen Linguistik* 20. 80-103.

Zaenen, Annie. 1997. Contrastive dislocation in Dutch and Icelandic. In Elena Anagnostpoulou, Henk van Riemsdijk & Frans Zwarts (eds.), *Materials on left dislocation*, 119-150. Amsterdam: John Benjamins.

CP and COMP in Diachrony

Nigel Vincent
The University of Manchester

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 314–333

http://csli-publications.stanford.edu/LFG/2019

Keywords: complementizers, accusative and infinitive, causatives, control, recomplementation, Latin, Romance

Vincent, Nigel. 2019. CP and COMP in Diachrony. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 314–333. Stanford, CA: CSLI Publications.

Abstract: This paper explores the ways in which a single Latin construction, the accusative and infinitive (AcI), has been replaced in different Romance languages. The parallel correspondence architecture of LFG provides an account which is more illuminating and theoretically more economical than that available to approaches which mediate all aspects of grammatical structure through a single set of syntactic categories and projections. Both categories and functions are seen to have their own diachronic profiles and the changes they exhibit over time do not necessarily proceed in parallel. More generally, the paper aims to show how both synchronic and diachronic data are relevant to the construction of theories about the structure and organization of human languages.

1. Introduction[1]

The domain of complement clauses is one where modern grammatical theories differ markedly both from traditional grammar and from each other. Consider the sentence in (1):

(1) Sarah believes that the train will be late.

Traditional grammar would label the string *that the train will be late* as noun clause object on the grounds that a) it could be replaced by a noun phrase such as *the rumour*, b) a noun phrase here would constitute the direct object of the verb *believe*, and c) the string *the train will be late* could stand as an independent finite clause with its embedded role here being signalled by the 'subordinating conjunction' *that*. Within a framework such as LFG with its distinction between f-structure and c-structure, debates have mainly centred around the object part of this traditional definition. Thus, in contrast to Bresnan & Kaplan's original proposal for a separate closed function COMP to be assigned to a constituent like *that the train will be late* Dalrymple & Lødrup (2005) argue that where a verb can take a direct nominal object, as *believe* does in English, this implies that the clausal complement should also have that function. COMP would then be reserved for verbs like *hope* which do not admit a nominal object. Others have gone a stage further and argued that COMP is redundant and all the functions of the embedded clause can be subsumed within OBJ or OBL (Alsina *et al* 2005). This debate is ongoing, with for example Patejuk & Przepiórkowski (2016) and Szűcs (2018) refining and providing further empirical evidence for the COMP-free approach while Belyaev *et al* (2017) argue for the continued recognition of a distinction between OBJ and COMP and the relevance of both in the domain of verbal complementation.

[1] This paper started life as a presentation at the teach-in on LFG and diachrony whch preceded LFG '19. I am grateful to the organizers, Wayan Arka and Jane Simpson, for inviting me to participate, to my co-presenters, Kersti Börjars and Louisa Sadler, to Mary Dalrymple, to those who attended and to the referees for their comments and suggestions. Responsibility for errors of fact or interpretation remains of course my own.

By contrast, there has been relatively little discussion within LFG of the categorial side of things. Although LFG is a framework which allows non-binary branching and exocentric configurations, both anathema to cartographic and nanosyntactic approaches, it is common to find the concept of CP carried over without comment from the Chomskyan tradition as the standard way to represent c-structures of constituents which begin with items like English *that,* French *que* and Hungarian *hogy*, and with it of course the implication that such items are heads. Yet such an assumption is by no means necessary. In the words of Pollard & Sag (1994: 44-5):

> "We are not claiming that the analysis of complementizers as heads is untenable, only that the fundamental intuition underlying such proposals raises as many questions as it answers … But if complementizers are not heads, then what are they? We will take the position that they are a subspecies of *marker*. On our account, a marker is a word that is 'functional' or 'grammatical' as opposed to substantive, in the sense that its semantic content is purely logical in nature (perhaps even vacuous). A marker, so-called because it formally *marks* the constituent in which it occurs, combines with another element that heads that constituent."

To this we may add the diachronic observation that the items that fall under the label of complementizer are always the product of processes of grammaticalization, and in that sense are different from lexical categories like noun and verb where core members may remain stable over centuries and even millennia. In other words, whatever is a C now will have been something else in the past.[2] And yet to date the debate has been exclusively based on synchronic evidence. In the present paper, therefore, we aim to introduce a diachronic dimension by means of a case study: the history of complementizers and complement clauses from Latin through to modern Romance. In section 2 we set out the Latin background before considering in sections 3 through 7 a variety of Romance developments and then in section 8 drawing some general conclusions.

2. Latin and the accusative and infinitive construction (AcI)

Latin had a variety of clausal complementation strategies but the central one for verbs whose semantics imply a propositional complement (thinking, saying, promising, hoping, knowing, believing, etc), and the one we will focus on here, was the one that goes by the traditional name accusative and infinitive construction (AcI) as exemplified in (2) - (4):

(2)	sese	confestim	supsequi	dicit
	REFL.3SG	immediately	follow.INF	say.PRS.3SG
	'He$_i$ says that he$_i$ will follow you immediately' (Caesar *Gall* 6.29.5)			

[2] Here and throughout, in order to avoid confusion, we will use C to refer to the category of complementizer and COMP to refer to the function, although the latter is also commonly used as a categorial label in the general literature.

(3) in aqua numquam credidi voluptatem
in water.ABL never believe.PERF.1SG pleasure.ACC.FSG
inesse tantam
in-be.INF such.ACC.FSG (Plautus *Rud.* 458)
'I have never believed that there was such pleasure in water'

(4) populus me vere iurasse iuravit
people.NOM.SG me.ACC truly swear.PERF.INF swear.PERF.3SG
'The people swore that I had sworn truly' (Cic *Fam* 5.2.7)

In these examples the governing verb takes a complement expressing the propositional content of the statement, belief or oath with the subject argument of the embedded verb in the accusative (*sese, tantam voluptatem, me*) and the verb in the infinitive, either present (*supsequi*, *inesse*) or perfective (*iurasse*). The accusative of the embedded subjects here must be generated clause internally since in many instances the governing verb either does not take a direct object, as with *dicere* 'say', or governs a different case, as with *credere* 'believe' which takes the dative (*crede.*IMP *mihi.*DAT 'believe me!'). Nor is the item in the accusative in semantic terms an argument of the governing verb. Note too that the embedded subject and the main clause subject can be coreferential as in (2) where the accusative of the AcI is the reflexive pronoun *sese*.

If we follow the account of this construction in Jøhndal (2012: 79-82), we can therefore assign these verbs the PRED values in (5):

(5) a. 'dicere <SUBJ, COMP>'
b. 'credere <SUBJ, COMP>'
c. 'iurare <SUBJ, COMP>'

In fact, however, nothing crucial hangs on assigning the second argument here the function COMP; the analysis would go through if we chose to follow Alsina *et al* (2005) and Patejuk & Przepiórkowski (2016) and assign it the function OBJ instead. Moreover, there are (admittedly rare) instances such as (6) in which an AcI (*iuraturas in feminae verba praetorias cohortis* 'that the praetorian cohorts would swear allegiance to a woman') can be co-ordinated with a simple NP (*consortium imperii* 'share of the power'):[3]

(6) quod consortium imperii iuraturas-que
that share.ACC power.GEN swear.FUTPRT.ACC.FPL-and
in feminae verba praetorias cohortis
in woman.GEN word.ACC.PL praetorian. ACC.FPL cohort. ACC.FPL

[3] To be precise, in example (6) the element *iuraturas*, marked with the co-ordinating affix *-que*, is the future participle of *iurare*, which taken together with the verb *esse* 'be' forms a future periphrasis. However in the AcI the auxiliary in its infinitival form is, as here, often omitted.

… speravisset
hope.PLUPERF.SUBJ.3SG
‘that she had hoped for a share in the empire and that the praetorian cohorts would swear allegiance to a woman’ (Tacitus *Ann* 14.11)

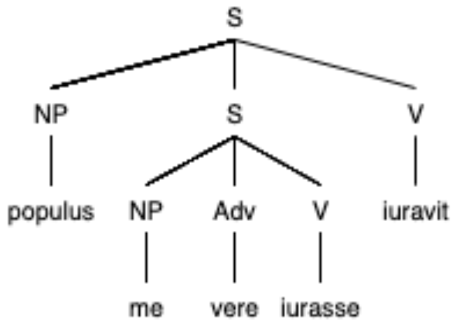
The pattern here is parallel to the English and Polish examples in (7) and (8) cited by Patejuk & Przepiórkowsk (2016) as a motivation for preferring OBJ to (X)COMP as the function to be assigned to clausal and infinitival complements (and see already Sag *et al* 1985):

(7) Pat remembered the appointment and that it was important to be on time.

(8) Nie chciał pić ani kanapki
NEG want.PST drink.INF nor sandwich.GEN
‘He didn’t want to drink nor (did he want) a sandwich’

What is key, however, is that an analysis of these patterns in terms of f-structure eliminates the need to postulate an empty complementizer head, so that the c-structure assigned for example to (4) would be as in (9):

(9)



Since the AcI, unlike a small clause, has the full range of tense, aspect and argument structure associated with a main clause, Minimalist or cartographic frameworks have little choice but to represent it as a CP. This is the analysis proposed, for example, in Oniga (2014: Ch 23) where the AcI is a CP with a zero complementizer which assigns accusative case, even though in general complementizers do not assign case and in this construction the C can never be overtly realized.

3. Complementizers and growing syntax in Romance

Central as the AcI is to the syntax of complementation in Latin, it comes over time to be replaced by a finite pattern introduced by items such as French *que* and Italian *che,* which derive from the Latin neuter relative pronoun *quid*. However, more frequently attested in Latin texts is the form *quod* which in origin had a causal value:

(10)

cum	tibi	agam		gratias
while	you.DAT	make.PRS.SUBJ.1SG		thanks.ACC
quod	me	vivere	coegisti	
because	me.ACC	live.INF	compel.PERF.2SG	(Cic *Att* 3.3.1)

‘while I may give thanks to you because you forced me to stay alive’

And in (11) we can see it being used to mark the complement of *credere* ‘believe’ in a text where the speaker is being identified as uneducated and uncultured and hence suggesting that this usage was part of popular speech at the time (2nd cent CE):

(11)

credo		nunc	quod	Pudentilla	me
believe.PRS.1SG		now	C	Pudentilla.NOM	me.ACC
in eo	tempore	non	amabat		
in that.ABL	time.ABL	NEG	love.IMPERF.3SG	(Apuleius *Apol* 79)	

‘I now believe that at that time Pudentilla did not love me’

It is *quod* which is the etymological source of the complementizer *co/cu* which survives in southern Italian dialects as in the Salentino example (12) and is already found in the earliest Italian text from 960 CE in (13):

(12)

oyyu	krai	ku	bbene	lu	Maryu
want.PRS.1SG	tomorrow	C	come.PRS.3SG	DEF	Mario

‘I want Mario to come tomorrow’

(13)

sao	ko	kelle	terre	… trenta	anni
know.PRS.1SG	C	that.FPL	land.FPL	30	year.PL
le	possette	parte	Sancti	Benedicti	
it.FPL	possess.PST.3SG	party	Saint	Benedict	

‘I know that those lands have belonged to the party of St Benedict for 30 years’

Interestingly, this complementizer is especially found, as in (12), with clauses that would take the subjunctive in those dialects which preserve that form, whereas causal clauses in Latin always take the indicative, thus suggesting a significant restructuring over time consistent with the loss of the inherent causal meaning.

By contrast, many southern dialects also have an indicative complementizer *ca/ka* which derives from another Latin causal marker *quia*, as in the Old Siclian example (14) from Rinaldi (2005: 473):

(14)

dicu	ka	dichi	beni
say.PRS.1SG	C	say.PRS.2SG	well

‘I say that you speak well’

This change must have started early since Bennett (1910: I,130) in his grammar of early Latin based on texts from the period before 100 BCE observes: “In apposition with neuter pronouns … the causal notion is usually very slight, *quia* having the force rather of ‘that’.”

Developments such as these — which we have only been able to sketch: see Ledgeway (2005) for a fuller treatment and further references — raise two questions of more general relevance in the present context. The first concerns the categorial status to assign to these items before they develop the functions exemplified here. One answer would be that even when they have semantic content of their own such as the causal meanings of Latin *quod* and *quia* they are nonetheless complementizers, so that the change is one involving semantic bleaching but not change of category. In this respect then the class of complementizers would be similar to prepositions, where it is common to recognise a distinction between items that have grammatical functions such as English *of* and French *de* and those with semantic content evidence by pairwise contrasts such as *before* vs *after* and *off* vs *on*. Taking this route would also provide a response to the observation by Pollard & Sag quoted above: only some complementizers would have a purely marking function but this would not stop them being treated as heads of CPs any more than it stops a constituent like *of my cousin* being defined as a PP in a construction like *proud of my cousin*. However, an argument against this view is provided by a 15th century Salentino example like (15) (cited in Ledgeway 2005: note 30):

(15) adivene perché ca Adamo lassao
happen.PRS.3SG because C A leave.PST.3SG
lo sua signo
DEF POSS.3SG sign
'it happens because Adam left his mark'

Here the complementizer *ca* co-occurs with the the word *perché* 'because'. Ledgeway's solution is to exploit the split CP hypothesis first put forward by Rizzi (1997) and developed extensively within the cartographic approach to clause stucture since that time. The item *ca* can then be assigned to the lowest functional head Fin while *perché* inhabits the specifier slot associated with the Interrogative head.

The second question follows on from the first, namely how are we to represent the mechanisms of change that are at work in these examples? Börjars *et al* (2016) argued that the development of grammaticalized definiteness markers in North Germanic was a case of 'growing syntax'. That is to say, rather than postulate a universal category DP with only some languages having an overt realization of D, it is proposed that the sole universal category is NP but that in some languages a D slot, and with it a DP projection, comes into existence over time via the well attested process of grammaticalization. By the same token, it might be suggested that there is only evidence for a CP in these structures once the C has emerged, once again via the mechanism of grammaticalization. The difference here, however, is that Latin did have CPs in other context as witness the items *ut* and *ne* in (16) and (17), which introduce subjunctive complements, respectively positive and negative, of the verbs *velle* 'want' and *timere* 'fear':

(16)	si	vis	ut	loquar
	if	want.PRS.2SG	C	speak.PRS.SUBJ.1SG

‘if you want that I should speak’ (Martial 5.52.6)

(17)	haec …	ne	impediantur	timeo
	these things	C-NEG	hinder.PRES.PASS.SUBJ.3PL	fear.PRS.1SG

‘I fear these things may be hindered’ (D Brutus 6 Cic *Fam*)

It seems then that in this case the CP has not so much ‘grown’ as ‘spread’. However, we defer further discussion of these issues to sections 6 and 7 below and turn instead to two other developments in Romance occasioned by the loss of the AcI.

4. Complex predicate formation: causative and perception verbs

Among the classes of verbs that in Latin could govern an AcI were causatives (18) and perception verbs (19):

(18)	ventus …	fecit …	spissescere	nubem
	wind.NOM.SG	make.PERF.3SG	thicken.INF	cloud.ACC.SG

‘the wind caused the cloud to thicken’ (Lucretius 6.176)

(19)	cum	illaec	autumare	illum	audio
	when	that.ACC.NEUT.PL	say.INF	that.ACC.M.SG	hear.PRS.1SG

‘when I hear that man say those things’ (Plautus *Am* 416)

In this instance, however, the diachronic development was not the replacement of the AcI by a finite clause but the fusing of the original two clauses into one through the formation of a complex predicate construction as in (20) and (21):

(20)	he	fet	veure	el problema	al director
	have.PRS.ISG	do.PSTPRT	see.INF	the problem	to.the director

‘I made the director see the problem’ (Catalan)

(21)	ho	udito	uscire	Paolo
	have.PRS.ISG	hear.PSTPRT	go out.INF	Paolo

‘I heard Paolo go out’ (Italian)

It is natural to assume that this reanalysis took place before the wholesale decline of the AcI pattern, with the consequence that these structures were not affected by the shift to clauses with overt complementizers described in the previous section.

A change like this, discussed in more detail in Börjars & Vincent 2017: 651-655), fits naturally within a framework like LFG since it is in essence a reorganization at the level of argument and f-structure, and can be handled directly in these terms rather than mediated through c-structure. The starting point is the representation for these verbs as in (22) and parallel to what we have already seen in (5) for *credere, dicere* and *iurare*:[4]

(22) a. 'facere <SUBJ, COMP/OBJ>'
b. 'audire <SUBJ, COMP/OBJ>'

The change then consists in the arguments of the embedded infinitival verb becoming dissociated from it and attaching instead to the light verb which heads the new complex predicate construction:[5]

(23) a. 'fare-V <SUBJ, OBJ, OBJ$_{\Theta}$>'
b. 'udire-V <SUBJ, OBJ, OBJ$_{\Theta}$>'

The new pattern is monoclausal whereas its historical antecedent was biclausal. Diagnostics for this changed state of affairs include first the fact that if the object is cliticised it must attach to the light verb and not to lexical verb of which it is a semantic argument. Thus, the clitic object version of (21) is *l'ho udito uscire* and not **ho uditolo uscire*. Second, if the lexical verb is unergative or unaccusative, the OBJ function of the complex predicate expresses that verb's semantic subject but if the lexical verb is transitive then its subject is forced to assume the OBJ$_{\Theta}$ role, hence *al director* in (20). In addition, monoclausal structures cannot be iterated. Contrast the grammaticality of iterated biclausal causatives in English examples such as *Bill made the director make his assistant answer the letter*.

5. Control verbs

So far, with complementizers we have seen developments that affect c-structure largely in isolation from f-structure while with complex predicate formation the essential shifts affect argument and f-structure, with any changes in syntactic constituency being consequential thereon. In this section, we will examine instead the changes which affect the control verb *velle* 'want', changes which concern both f- and c-structure. Once again the starting point is the AcI as in (24):

(24)	volo	te	uxorem	domum	ducere
	want.PRS.1SG	you.ACC	wife.ACC	home.ACC	lead.INF

'I want you to take a wife' (Plautus *Aul* 149)

[4] We use the notation COMP/OBJ to indicate that nothing hangs on the choice between either the COMP-based account of the OBJ one, although the fact that both *facere* 'do' and *audire* 'hear' can also occur with simple nominal objects suggests that the OBJ-based analysis might be preferable.

[5] We use Italian for exemplificatory purposes here but the same would hold for other Romance reflexes of these verbs such as French *faire,* Spanish *hacer, oír* and indeed for cases in which the lexical realization of the light verb component of the construction has changed as with French *entendre* 'hear' (< Latin *intendere* 'stretch, direct attention to') or Portuguese *mandar* 'make' (< Latin *mandare* 'send'). For further discussion of the argument assigning mechanisms involved here see Alsina (1996) and Butt (2010).

At the same time we have also seen that this verb may take a finite CP complement, as in (16). Both these examples involve different subjects in the main and embedded clauses, while in the same subject construction the most common pattern is a simple infinitive as in (25):

(25)	potare	ego	hodie,	Euclio,	tecum	volo
	drink.INF	I	today	Euclius	you-with	want.PRS.1SG

‘I want to drink with you today, Euclius’ (Plautus *Aul* 569)

In addition, in the words of Jøhndal (2012: 92), ‘surprisingly, we also find the AcI under coreference’ as in (26) (= his 110), though as he goes on to note examples of this type are less frequent that the more usual plain infinitive in the same subject construction:

(26)	volo	me	placere	Philolachi
	want.PRS.1SG	me.ACC	please.INF	Philolaches.DAT

‘I want to please Philolaches’ (Plautus *Mos* 167)

Given what we have seen so far, it is less surprising that the AcI disappears in both its same and different subject variants, leaving a pattern of alternation between a bare infinitive and a finite CP as in the Italian examples in (27):

(27)	a.	voglio	partire	domani		
		want.PRS.1SG	leave.INF	tomorrow		
		‘I want to leave tomorrow’				
	b.	voglio	che	tu	parta	domani
		want.PRS.1SG	C	you.NOM	leave.PRS.SUBJ.2SG	tomorrow
		‘I want you to leave tomorrow’				

What is less expected is that the same subject variant in (27a) has both a monoclausal and a biclausal version, as can be seen from the alternative positions of the clitic *ci* ‘there’ in (28):

(28)	a.	voglio	andarci	
		want.PRS.1SG	go.INF-there	
		‘I want to go there’		
	b.	ci	voglio	andare
		there	want.PRS.1SG	go.INF

(28a) and (28b) are synonymous but (28b) has undergone so-called ‘restructuring’ to become a single clause as further evidenced by the fact that in the periphastic perfect *volere* requires the auxiliary *essere* ‘be’ appropriate to *andare* ‘go’ rather than *avere* ‘have’ which it requires in isolation: hence *ci sono voluto andare* ‘I wanted to go there’ and not **ci ho voluto andare* but *ho voluto una birra* ‘I wanted a beer’.

How then are we to model the diachronic trajectory here? As we have seen in (5), Jøhndal (2012) proposes to assign the function COMP to AcI across the board, which has the effect of treating both (24) and (26) as instances of anaphoric control. By contrast, he proposes to treat the bare infinitive type (25) as an instance of functional control with the infinitival predicate being assigned the function XCOMP. The loss of the AcI has the effect, then, of creating a clear alternation between different subject/anaphoric control and same subject/functional control. This closely parallels the same diachronic sequence postulated in the account in Börjars & Vincent (2019) of the mechanisms underlying the development of *wil-, the Germanic cognate of Latin *velle*, into the modern English auxiliary *will* via the 'want' meanings seen in Old English *willan* and modern Swedish *vilja*, with functional control seen as the intermediate stage between anaphoric control (and more particularly 'quasi-anaphoric control' in the sense of Haug 2013) and the PRED-free tense/aspect value of English *will*. More generally, such accounts provide a natural way of modelling the kind of semantic 'bleaching' standardly associated with the process of grammaticalization.

In summary, then, the history of the Latin 'want' verb *velle* and its Romance descendants French *vouloir* and Italian *volere* provides evidence of two distinct diachronic trajectories, one having to do with f-structure and one with c-structure.[6] Such developments can be easily accommodated within a framework like LFG but create an analytical problem for cartographic approaches, where a decision has to be made as to whether to accord an item like *volere* the status of an independent main verb or to treat it as a functional head. Grano (2015: 89) opts for the latter solution: 'Following Cinque (2004), I take the cross-linguistically robust restructuring of *want* as decisive in classifying *want* as a functional head in the inflectional layer of the clause.' The problem then is that as such it cannot govern a CP. To accommodate examples like (27b) Grano is obliged to postulate an intervening silent HAVE as the lexical head of 'want' clauses plus the further assumption that the complement of HAVE is not a CP but a *v*P. This in turn requires him to deny complementizer status to the Italian item *che* in (27b) despite the fact that *che* has a standard complementizing function in clauses dependent on verbs of saying, thinking and the like (Grano 2015: 83, note1). If, on the other hand, he had chosen to classify *want* as a lexical V head, his framework has no obvious way to handle the alternation seen in (28) without a further set of arbitrary assumptions. There is not space here to go into a detailed analysis of these proposals, but it suffices to note the problems that arise within an approach in which one variant has to be given derivational priority over the other, problems that disappear in a model such as LFG where differing c-structures can be mapped onto the same f-structure.

Further evidence of the way c-structure and f-structure may develop independently is to be seen in southern Italian dialects such as Salentino (Calabrese 1993), where we find the patterns in (29):

---

[6] There are similar patterns to be seen in other Romance languages such as Spanish, Portuguese and Sardinian, but with the additional complication that the relevant lexical item is not a reflex of *velle* but the result of a lexical semantic shift of Latin *quaerere* 'seek' to yield Spanish/Portuguese/Galician *querer* and Sardinian *kerrere.*

(29) a. voggyu lu kattu
want.PRS.1SG it buy. PRS.1SG
b. lu voggyu kattu
it want.PRS.1SG buy. PRS.1SG
c. voggyu ku lu kattu
want.PRS.1SG C it buy. PRS.1SG
I want to buy it'
d. *lu voggyu ku kattu

The difference here is that the complement of the 'want' verb in these dialects is expressed by a finite form rather than the infinitive even with the same subject construction. Restructuring is still possible, however, as evidenced by the equivalence of (29a) and (29b), but if the complementizer *cu* (<Lat *quod*) is present as in (29c) then restructuring is blocked and hence the ungrammaticality of (29d).

Conversely, elsewhere in southern Italy and in Sardinia it is the infinitive which generalises leading to patterns like old Sicilian (30) and modern Sardinian (31):

(30) a. eu nun vi voglu veniri
I NEG there want.PRS.1SG come.INF
'I do not want to come there' (Rinaldi 2005: 152)

b. vulissi homu tu non chi essiri
want.PST.SUBJ.3SG one you.NOM NEG there be.INF
'one would like you not to be there' (Bentley 2014: 99)

(31) a. non kèlio vénnere
NEG want.PRS.1SG come.INF
'I do not want to come'

b. non kèlio a vénnere tue
NEG want.PRS.1SG C come.INF you.NOM
'I do not want you to come' (Jones 1992)

It is to be noted here that in different ways the monoclausal/biclausal distinction is still evident: in (30a) the clitic precedes the 'want' verb while in (31a) there is no complementizer in contrast to the presence of *a* in (31b). Note too that in both (30b) and (31b) the subject of the infinitive is in the nominative, thus marking this out as a Romance development rather than a continuation of the Latin AcI.[7] Thus, once again parallel argument structures map onto different grammatical categories.

[7] Note that the nominative plus infinitive construction here is different from the one that goes by that name in Latin. The latter is simply a passivized variant of the AcI: *Marcus*.NOM *abire*.INF *dicitur*.PRES.PASS.3SG 'Marcus is said to be leaving' (Jøhndal 2012: 61).

6. Prepositional complementizers and split CP

What we have seen in the case of the Latin and Romance 'want' verbs is on the one hand the replacement of the AcI by a finite complement clause when the subjects differ and the continuity of the bare infinitive construction as the only option when the subjects of the main and complement clauses coincide. This, however, is a combination of properties virtually unique to 'want'. The more common situation is the kind of alternation seen in the French examples (32) and (33), where the dependent infinitive is introduced by a marker of its own such as *de* or *à*:

(32) a.

J'ai	décidé	de	partir	demain
I-have.PRS.1SG	decide.PSTPRT	DE	leave.INF	tomorrow

'I have decided to leave tomorrow'

b.

J'ai	décidé	qu'on	partira	demain
I-have.PRS.1SG	decide.PSTPRT	C-one	leave.FUT.3SG	tomorrow

'I have decided that we will leave tomorrow'

(33)

Pierre	m'a	invité	à	venir	demain
Pierre	me-have.PRS.3SG	invite.PSTPRT	A	come.INF	tomorrow

'Pierre has invited me to come tomorrow'

The issue then is how to model these items. Etymologically there is no doubt that we are dealing with reflexes of the Latin prepositions *ad* 'to' and *de* 'from', which can also be seen in expressions like *à Paris* 'to/in Paris' and *de Londres* 'from/of London'. The complication is that in Latin prepositions do not co-occur with infinitives so the functions exemplified in (32) and (33) are Romance innovations, where they serve as non-finite alternants of items like *que* and hence the label 'prepositional complementizer' which they have acquired in the literature. Within LFG the choice lies between treating them as prepositions that take infinitival complements, thus yielding structures like (34a), or as complementizers as in (34b):

(34) a. [$_{PP}$ [$_{P}$ de] [$_{VP}$ partir demain]], [$_{PP}$ [$_{P}$ à] [$_{VP}$ venir demain]]

b. [$_{CP}$ [$_{C}$ de] [$_{VP}$ partir demain]], [$_{CP}$ [$_{C}$ à] [$_{VP}$ venir demain]]

There are arguments in favour of both. Formal identity might lead one to prefer the prepositional solution, whereas the pattern of finite/non-finute alternation seen in (35) argues in favour of the complementizer account:

(35) a. *avant de partir demain* 'before leaving tomorrow'

b. *avant que tu partes demain* 'before you leave tomorrow'

Alternatively one can seek to import solutions developed within other frameworks. Thus, Abeillé *et al* (2006) introduce the concept of a 'weak head' for precisely these cases, where a weak head is characterised as having the status of a 'prep-word', that is to say the same as a regular preposition, but it is weak in the sense that it yields its head value to the item with which it co-occurs so that overall the structure is for selection purposes headed by the V and not by the P. In LFG terms, this is very similar to the role played by non-projecting items (see Vincent & Börjars 2019 for further discussion).

A different approach is that adopted within the cartographic model of the left periphery proposed by Rizzi (1997) and alluded to above. Rizzi notes that whereas *que* and the following material in an example like (35b) can be separated by fronted of parenthetical elements, the sequence *de partir* in (35a) can only be separated by verbal clitics as in *d'en partir* 'leave from there' or *d'y aller* 'go to there'. He therefore proposes to break C down into a series of hierarchically organized functional heads, with a finite complementizer like *que* occupying the highest head, labelled Force, while items like *à* and *de* occupy the lowest head, labelled Fin. In other words, in a string like *avant de partir*, there would be a full lexical preposition *avant* 'before' linked to an infinitive by a complementizing particle *de*, which has here lost its etymological status as a preposition, so that the structure is similar to that discussed for the string *perché ka* 'because that' in example (15). In general, LFG has avoided the proliferation of functional heads that is characteristic of the cartographic approach but, as the data from this section and the following one suggest, this may be one instance where the price is worth paying.

7. Recomplementation

The phenomena we have considered so far play to LFG's strengths insofar as they involve patterns of interaction between different levels of structure with no ontological or derivational priority being given to one type of structure above all others. In particular, there is no central role for categorial syntactic representation. We turn our attention now to something which comes with the historical development of complementizers and which at first sight looks to argue strongly for a configurational account, namely complementizer doubling or what in the recent literature has come to be so-called recomplementation. This is the phenomenon whereby complementizers are repeated around a fronted element as in the English examples in (36) - (38):

(36) The party opposite said [**that** if we cut 6 billion from the budget, **that** it would end in a catastrophe] [David Cameron, Prime Minister's Questions in the UK House of Commons]

(37) I'm glad [**that,** whoever talked Strauss into it, **that** they did [Geoff Boycott, BBC Radio 5]

(38) 'Forster once wrote **that** if he had to choose between betraying his friends or his country, **that** he hoped he would have the courage to betray his country.'
[Christopher Catherwood *The Cuckoos' Nest. Five Hundred Years of Cambridge Spies*, Cambridge, Oleander Press, 2013, p.59]

One immediate response to cases such as these might be to dismiss them as errors. Examples like (36) and (37) are drawn from off-the-cuff spoken language — a response to a question in Parliament in the case of (36) and sports commentary in (37) — a genre which inevitably involves hesitations and repetitions that go beyond the bounds of grammar. And occasional written examples such as (38) might be challenged as oversights that more careful proofreading could have eliminated.

However, in Romance at least, the phenomenon is too richly attested in a range of modern and medieval varieties for this to be a convincing escape route. Thus, when discussing the evidence from early Italian dialects, Ledgeway (2005: 3008) observes: ‘The examples to be considered, though not so numerous in each single text as to be legitimately considered a core grammatical phenomenon, do however occur in sufficient number and across a wide range of texts from different regions to be interpreted as the reflex of a regular structural phenomenon’. The fact too that our example (40) is from a canonical literary figure like Boccaccio makes it hard to dismiss such cases out of hand. Representative instances then are given here in (39) from modern Portuguese and (40) from old Italian:[8]

<table>
<tr><td>(39)</td><td>Duvido</td><td>que</td><td>a</td><td>Ana</td><td>que</td><td>goste</td><td>de ópera</td></tr>
<tr><td></td><td>doubt.PRS.1SG</td><td>C</td><td>to</td><td>Ana</td><td>C</td><td>please.PRS.SUBJ.3SG</td><td>of opera</td></tr>
<tr><td></td><td colspan="5">‘I doubt whether Ana likes opera’</td><td colspan="2">(Mascarenhas 2014)</td></tr>
</table>

<table>
<tr><td>(40)</td><td>ti</td><td>priego</td><td>che,</td><td>se</td><td>egli</td><td>avviene</td><td></td><td></td><td></td></tr>
<tr><td></td><td>you</td><td>beg.PRS.1SG</td><td>C</td><td>if</td><td>it</td><td>happen.PRS.3SG</td><td></td><td></td><td></td></tr>
<tr><td></td><td>ch’io</td><td>muoja</td><td>che</td><td></td><td>le</td><td>mie</td><td>cose</td><td>ed</td><td>ella</td></tr>
<tr><td></td><td>C I</td><td>die.PRS.SUBJ.1SG</td><td>C</td><td></td><td>the</td><td>my</td><td>things</td><td>and</td><td>her</td></tr>
<tr><td></td><td>ti</td><td>sieno</td><td colspan="4">raccomandate</td><td></td><td></td><td></td></tr>
<tr><td></td><td>you</td><td>be.PRS.SUBJ.3SG</td><td colspan="4">entrusted.PSTPRT</td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="9">‘I beg you that, if I die, my things and her should be entrusted to you’</td></tr>
<tr><td></td><td colspan="9">(Boccaccio Decameron 2,7)</td></tr>
</table>

In (39) the complementizer *que* appears at the beginning of the complement clause and then again after the fronted topicalized item *a Ana*. In (40) we see a similar pattern, with the difference that the two complementizers appear before respectively the protasis and the apodosis of the embedded conditional sentence. On the usual assumption that the protasis of a conditional sentence is a kind of topic these examples can be made to fit very neatly into Rizzi’s extended left periphery with the first occurrence of the complementizer in the Force head and the second in the Fin head and with the topicalized element as specifier of the Topic head:

(41) [$_{\text{Force P}}$ [$_{\text{Force}}$ que] [$_{\text{TopP}}$ [$_{\text{PP}}$ a Ana] [$_{\text{Top}}$ Ø]] [$_{\text{FinP}}$ [$_{\text{Fin}}$ que]]]

[8] Further examples and references to the literature can be found in Wanner (1995), Paoli (2003, 2007) and Munaro (2016). Salvesen & Walkden (2017) also cite examples from Old English.

Indeed, just such an analysis is proposed by Ledgeway (2005) and Villa-Garcia (2015), while Radford (2018) develops a similar account for the spoken English examples (37) and (38).

The challenge for LFG is then to see how this kind of data can be accommodated. One possibility might be to treat this as CP recursion with the fronted element located in the specifier position of the higher C:

(42) [$_{CP}$ [$_{C}$ que] [$_{TopP}$ [$_{PP}$ a Ana] [$_{Top}$ Ø]] [$_{CP}$ [$_{C}$ que]]] …

However, this either implies implausibly that a complementizer can take a CP as its own complement or it is simply a notational variant of the split CP analysis. In this connection, it is instructive that Zipf & Quaglia (2017) propose an LFG-analysis of a different Italian phenomenon which includes what they call a C-structure template (their Figure 7) akin to (42) and then comment in a footnote (p.399, note 6) that this ‘is not meant to represent the case of CP recursion but rather two different C-related projections’ adding that nonetheless they do not adopt Rizzi’s labels but ‘prefer remaining neutral to the specific implications of these projections’. It is hard, however, to see what ‘remaining neutral’ in these circumstances can mean; it would appear that *de facto* if not *de nomine* they have incorporated the concept of split CP into the range of phrase structures permitted within LFG. Nor is there anything inherently implausible about such a conclusion. There is no universally fixed limit to the range of c-structure categories that natural language data require to be recognised if they are to be analysed in proper detail. The question is rather whether a categorial analysis is the right solution for any subset of such data. In the present context it is hard to avoid the conclusion that this is indeed the right solution, and that therefore a c-structure with a split CP will need to be deployed even within a framework such as LFG.

At the same time it is easy to understand the reluctance to go down this route since it can lead to the explosion of functional heads that is characteristic of recent nanosyntactic work (see for example Baunaz 2018 and references there). An alternative therefore would be to follow the idea of Sag & Pollard and abandon the idea of a complementizer as a head and treat it as simply a structural marker that can be inserted as pragmatic circumstances dictate. Such an account would be consistent with the occasional attested instances of complementizer tripling as in the old Neapolitan (43), taken from a letter dated 1353, where the function of the complementizers appears to be to break the text down into rhythmical or rhetorical chunks:[9]

9 In order to facilitate legibility I have deliberately not glossed this example but hopefully the literal translation plus the complementizers in bold will make the intended structure clear.

(43) Pregove, madama, per l'amor di Dio, **che** de chilli dinare che eo agio vostri **che** si non vi fusse troppo sconço **che** mi 'ndi impristiti una unça.
'I beg you, lady, for the love of God, **that** of that money that I have of yours **that** if it wasn't too much trouble **that** you should lend me some.

For the present we leave open the question as to which these two analytical routes it is preferable to follow. Either way the possibility within LFG of consigning the pragmatic interpretation of the fronted elements to an independent dimension of i-structure (Dalrymple & Nikolaeva 2011) means that the number of slots in the c-structure can be kept to a minimum.

8. Conclusion

In summary, then, what the present paper has sought to do is explore the various ways in which a single Latin construction, the AcI, has been replaced with different structures in different Romance languages and in a range of different syntactic contexts. In the course of the analysis we have seen how the parallel correspondence architecture of LFG, with its separation of a-, f-, c- and i-structure, provides an account which is both more illuminating and theoretically more economical than that available to approaches which mediate all aspects of grammatical structure — and therefore all aspects of change — through a single set of syntactic categories and projections. We have demonstrated that both categories and functions have their own diachronic profiles and that the changes they exhibit over time do not necessarily proceed in synch with each other. At the same time we have raised some questions about the precise nature of such categories and whether for example Rizzi's split CP model needs to be incorporated into LFG, in particular as a way of dealing with the phenomenon of recomplementation. More generally, our work has been inspired by the conviction that any theory or framework needs to be able to accommodate both synchronic and diachronic data and that there is no reason to privilege one over the other if the aim is to understand the mechanisms and processes at work in the organization of human language.

References

Abeillé, Anne, Olivier Bonami, Danièle Godard & Jesse Tseng. 2006. The syntax of French *à* and *de*: an HPSG analysis. In Patrick Saint-Dizier (ed.) *Dimensions of the Syntax and Semantics of Prepositions*. 147-162. Dordrecht: Springer.

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar. Evidence from Romance*. Stanford, Ca: CSLI.

Alsina, Alex, Tara Mohanan & K. P. Mohanan. 2005. How to get rid of the COMP. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2005 Conference,* 83-103. Stanford: CSLI, http://csli-publications.stanford.edu/

Baunaz, Lena. 2017. Decomposing complementizers: the functional sequence of French, Modern Greek, Serbo-Croatian, and Bulgarian complementizers. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds) *Exploring Naonsyntax.* 149-179. Oxford: Oxford University Press.

Belyaev, Olga, Anastasia Kozhemyakina & Natalia Serdobolskaya. 2017. In defense of COMP: complementation in Moksha Mordvin. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2017 Conference,* 83-103. Stanford: CSLI, http://csli-publications.stanford.edu/

Bennett, Charles E. 1910. *Syntax of Early Latin.* Boston: Allyn and Bacon.

Bentley, Delia. 2014. On the personal infinitive in Sicilian. In Paola Benincà, Adam Ledgeway, Nigel Vincent (eds) *Diachrony and dialects. Grammatical change in the dialects of Italy,* 96-115. Oxford: Oxford University Press.

Börjars, Kersti, Pauline Harries & Nigel Vincent. 2016. Growing syntax: the development of a DP in North Germanic. *Language* 92: e1-e37.

Börjars, Kersti & Nigel Vincent. 2017. Lexical-Functional Grammar. In Adam Ledgeway & Ian Roberts (eds) *The Cambridge Handbook of Historical Syntax*, 642-663. Cambridge: Cambridge University Press.

Börjars, Kersti & Nigel Vincent. 2019. Modelling step change: the history of WILL-verbs in Germanic. In Nuria Yañez-Bouza, Willem Hollmann, Emma Moore & Linda van Bergen (eds) *Categories, Constructions and Change in English Syntax*. 283-314. Cambridge: Cambridge University Press.

Butt, Miriam. 2010. The light verb jungle: still hacking away. In Mengistu Amberber, Brett Baker & Mark Harvey (eds) *Complex predicates. Cross-linguistic perspectives on event structure*. 48-78. Cambridge: Cambridge University Press.

Calabrese, Andrea. 1993. The sentential complementation of Salentino: a study of a language without infinitival clauses. In Adriana Belletti (ed.) *Syntactic Theory and the Dialects of Italy.* 28-98. Turin: Rosenberg & Sellier.

Cinque, Guglielmo. 2004. 'Restructuring' and functional structure. In Adriana Belletti (ed.) *Structures and Beyond: The Cartography of Syntactic Structures.* 132-191. Oxford: Oxford University Press,

Dalrymple, Mary & Helge Lødrup. 2000. The grammatical functions of complement clauses. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2000 Conference.* Stanford: CSLI, http://csli-publications.stanford.edu/

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure.* Cambridge: Cambridge University Press.

Grano, Thomas. 2015. *Control and Restructuring.* Oxford: Oxford University Press.

Haug, Dag. 2013. Partial control and anaphoric control in LFG. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2013 Conference*, 274-294. Stanford: CSLI, http://csli-publications.stanford.edu/

Jøhndal, Marius L. 2012. *Non-finiteness in Latin.* PhD dissertation, University of Cambridge.

Jones, Michael. 1992. *Sardinian Syntax*. London: Routledge.

Ledgeway, Adam. 2005. Moving through the left periphery: the dual complementiser system in the dialects of southern Italy. *Transactions of the Philological Society* 103: 339-396.
Mascarenhas, Salvador. 2014. Complementizer doubling in European Portuguese. *Rivista di Grammatica Generativa* 36. 105-116.
Munaro, Nicola. 2016. A diachronic approach to complementizer doubling in Italo-Romance and the notion of downward reanalysis. *Rivista di Grammatica Generativa* 38:215-228.
Oniga, Renato. 2014. *Latin: A Linguistic Introduction*. Oxford: Oxford University Press.
Paoli, Sandra. 2003. *COMP and the left periphery: comparative evidence from Romance*. PhD dissertation, Department of Linguistics, University of Manchester.
Paoli, Sandra. 2007. The fine structure of the left periphery: COMPs and subjects. Evidence from Romance. *Lingua* 117:1057-79.
Patejuk, Agnieszka & Przepiórkowski, Adam. 2016. Reducing grammatical functions in LFG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds) *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 541-559. Stanford: CSLI, http://csli-publications.stanford.edu/
Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar.* Chicago: University of Chicago Press.
Radford, Andrew. 2018. *Colloquial English. Structure and variation.* Cambridge: Cambridge University Press.
Rinaldi, Gaetana Maria. 2005. *Testi d'archivio del Trecento*. Palermo: Centro di studi filologici e linguistici siciliani, (Collezione di testi siciliani dei secoli XIV e XV, 24-25).
Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.) *Elements of Grammar.* 281-337. Dordrecht: Kluwer.
Sag, Ivan, Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Co-ordination and how to distinguish categories. *Natural Language and Linguistic Theory* 3: 117-171.
Salvesen, Chrisine & George Walkden. 2017. Diagnosing embedded V2 in Old English and Old French. In Eric Mathieu & Robert Truswell (eds) *Micro-change and Macro-change in Diachronic Syntax*. 168-181. Oxford: Oxford University Press.
Szűcs, Péter. 2018. A COMP-less approach to Hungarian complement clauses. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2018 Conference,* 325-342. Stanford: CSLI, http://csli-publications.stanford.edu/
Villa-Garcia, Julio. 2015. *The Syntax of Multiple* que *Sentences in Spanish*. Amsterdam: John Benjamins.
Vincent, Nigel & Kersti Börjars. 2010. Grammaticalization and models of language. In Elizabeth Traugott & Graeme Trousdale (eds) *Gradience, gradualness and grammaticalization,* 279-297. Amsterdam: Benjamins.
Vincent, Nigel & Kersti Börjars. 2019. Heads and history. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds) *Syntactic Architecture and its Consequences: Synchronic and Diachronic*

Perspectives. Vol 1: *Syntax inside the Grammar*. 135-159. Berlin: Language Science Press.

Wanner, Dieter. 1995. Les subordonnées à double complémentateur en roman mediéval. In Giovanni Ruffino *et al* (eds) *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza, Palermo, 18-24 settembre 1995*. Tübingen: Niemeyer, Vol 1, pp. 421-433.

Zipf, Jessica & Stefano Quaglia. 2017. Asymmetries in Italian matrix *wh*-questions: word order and information structure. In Miriam Butt & Tracy Holloway King (eds) *Proceedings of the LFG 2017 Conference,* 387-405. Stanford: CSLI, http://csli-publications.stanford.edu/

Morphology or syntax: The two types of non-agreeing verb

Glenn Windschuttel

University of Newcastle, Australia

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 334–352

http://csli-publications.stanford.edu/LFG/2019

Keywords: agreement, sporadic agreement, non-agreeing, morphology, syntax, objects

Windschuttel, Glenn. 2019. Morphology or syntax: The two types of non-agreeing verb. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 334–352. Stanford, CA: CSLI Publications.

Sporadic verb agreement, that is, where some verbs show agreement and others do not, is not a common feature of languages around the world, especially if lexical (Fedden 2019, Windschuttel 2019a). Where it affects objects, there are two types. In the first type, there are other syntactic differences between the verbs and their objects, not just agreement. Dahlstrom (2009) analysed this as a difference in the grammatical functions they subcategorise for, OBJ where indexed and OBJ_{θ} where unindexed. The other type cannot be reconciled to this analysis, the difference in agreement behaviour having no wider syntactic significance. Instead, morphology is the only difference. These two types, morphological and syntactic, parallel the distinction between morphological and syntactic ergativity both in behaviour and analysis.

1 Introduction[1]

Sporadic agreement was coined by Corbett (2006:17) to describe the situation where agreement only appears on a proper subset of the target wordclass.[2] Very little has been written about this wider phenomenon under this name (a search reveals only a number of conference presentations by Fedden 2017, 2017a, 2017b and a chapter, Fedden 2019). Nonetheless, research into agreeing and non-agreeing verbs precedes this term. The object agreeing class in the Trans-New-Guinea (TNG) languages, for example—defined in contrast to their non-agreeing transitive counterparts—were noticed as early as Pilhofer (1933: he called them "Objektverben", object verbs, see also Suter 2012, Foley 1984, 2000, Windschuttel 2017, 2019a) while the sporadic nature of sign language agreement has long been recognised (Meier 1982, Padden 1988; if it is actually agreement, see discussion in the conclusion).

Common examples of sporadic agreement include uninflected adjectives such as German *lila* and *rosa* which appear bare in attributive position such as *ein rosa Kleid* (Spencer 2009:209; cf: a regular adjective *blau* in the same phrase, *ein blaues Kleid*). These adjectives do not bear the regular

[1] I would like to thank all those present at the poster session, the proceedings editors, reviewers, Guillaume Jacques and Sebastian Fedden. Acknowledgement must go to the Australian Government Research Training Program (RTP) who funded some of this research. Finally, I thank the Kui community in Buraga, Lerabain and Moru for their hospitality and assistance.

[2] This is itself a subtype of 'sporadic inflection', for example, English s*heep* with respect to number inflection (M. Baerman p. c. in Fedden 2019). Clearly, this concept is also related to uninflectability.

agreement morphology that expresses the gender, case and number of the head noun.

Such examples are quite different from the focus of this paper since they are completely uninflecting, not admitting any inflection, not just agreement. Moreover, it is possibly the final /a/, unusual for German phonology, that explains the missing agreement. Contrastingly, the examples in this paper will be not so simply explained.

Agreement need not be completely absent. It may be that only a certain type of agreement is missing on the sporadic items.[3] This is the common pattern in the TNG object verb languages where prefixal object agreement is sporadic while suffixal subject agreement is not.

Tairora in the TNG subfamily, Kainantu-Goroka, provides an example. Object verbs like *aaru* 'hit' are prefixed directly to reference their objects, as is exemplified below in (1). The prefix *h-* on the verb indexes the first singular object. For non-prefixing verbs like *tave* 'see', this is impossible: for example, in (2), *tave* has a first singular object but no prefix. Nonetheless, both groups of verbs index the subject with a suffix as both these examples show.[4]

	TAIRORA		
(1)	Aaqu	ti	h-aaru-antora.
	rain	1SG.OBJ	1SG.OBJ-hit-3SG.AVOL
	'I don't want the rain to hit me.' (Vincent 2003:599)		

(2)	Ti	tave-ro.
	1SG.OBJ	see-3SG.PST
	'He saw me.' (Vincent 2003:584)[5]	

3 This could be called relative sporadic agreement following Windschuttel (2018) on uninflectedness (this could even be subsumed under uninflectedness but relative to object agreement, etc.).

4 There exists wide variation how objects are expressed with non-agreeing verbs (Windschuttel 2019a). Typically, full NPs are accepted, at least, where they are singular and inanimate or obviative. In other cases, some languages use free pronouns as Tairora here exemplifies. In other TNG languages, an agreeing verb coocurrs acting as an auxiliary carrying the agreement information (Foley 1984, Windschuttel 2019a) while in the Algonquian language, Plains Cree, there is no way of expressing objects of other persons with these verbs (Tollan & Oxford 2018).

5 Non Leipzig glosses: AUTO=autobenefactive, FACT=factual, IFR=inferential, N.PST=non-past, MED=medial, PART=partitive, AVOL=avolitional

Another example is found in many Algic languages with ‘pseudotransitive’ verbs (also known as VAIO, animate intransitive verbs with object).[6] The object is ignored by agreement while the subject continues to be indexed. This pattern in the Algic language, Meskwaki, has already been given an LFG analysis, Dahlstrom (2009), the difference between the agreeing and non-agreeing verbs being the syntax of the object. Agreeing verbs subcategorise for OBJ and non-agreeing verbs $OBJ_θ$. The next section will apply this analysis to the Tibeto-Burman language, Japhug, which acts similarly.

However, this analysis does not hold for all examples of non-agreeing verbs. In the Papuan language, Kui, the absence of agreement does not correlate with any syntactic difference in the objects. Instead, the morphology of the verb looks to be the only difference as Section 3 will detail. There are a number of possible analyses for this in LFG depending the morphological theory chosen. Two are given in Section 4. Whatever the details of its analysis, this produces two types of sporadic agreement according to whether the classes are syntactic or only morphological.

2 Syntax explains the absence of agreement

Verbal agreement is sensitive to both arguments in normal transitive clauses in Japhug (Rgyalrong in Tibeto-Burman). A notable exception are a small class that do not index their objects, the semi-transitives. There is a relatively simple explanation for the behaviour of this class: they subcategorise for $OBJ_θ$. Other features of the syntax of these objects and $OBJ_θ$ in this language support this, not only the absence of agreement. This is the analysis Dahlstrom (2009; based largely on the Relational Grammar account of Rhodes 1990) gave to the pseudotransitive verbs in the Algic language, Meskwaki (a typological connection between the two was recognised by Jacques 2016).

Japhug has ergative alignment, the transitive subject marked by the ergative postposition *kɯ* (Jacques 2016).[7] This can be seen in (3). By contrast, the object is unmarked. This is just like the subject of an intransitive verb as displayed in (4).

[6] Relative sporadic agreement is also found in those Nakh-Daghestanian languages with person agreement such as Dargwa, where person suffixes appear on all verbs while only some verbs take gender-number prefixes indexing their absolutive argument (Belyaev 2013).
[7] Relativisation which groups A and S together shows that ergativity in Japhug is only a surface phenomenon (Jacques 2016).

JAPHUG

(3) ...ɯʑo kɯ qɤjɣi χsɯm lo-βzu.
3SG ERG loaf three IFR-make
'...she made three loaves.' (Jacques 2004:444)

(4) Tɤ-tɕɯ nɯ jo-ɕe.
INDEF.POSS-boy DEM IFR-go
'The boy went (there).' (Jacques 2016:2)

Agreement in Japhug is aligned hierarchically, sensitive to both arguments according to an inverse system (Jacques 2010). In (5) we see a direct sentence, with a suffix indexing the second person subject. In (6) the situation is reversed, with a third singular acting on a second person object; however, the second person suffix is the same but the role it indexes is changed by the inverse prefix. Japhug makes extensive use of zero anaphora and a single verb can form a complete utterance as in both of these examples (Jacques 2010; the agreement itself may also be pronominal).

JAPHUG

(5) Pɯ-tɯ-mtó-t.
AOR-2-see-PST
'You saw him/her/it.' (Jacques 2010:129)

(6) ...βdɯt kɯ tɯ́-wɣ-ndza.
demon ERG 2-INV-eat:FACT
'...the demon will eat you.' (Jacques 2014a:309)

Alongside the basic transitive pattern exemplified above, there is a class of two-place verbs, the semi-transitives, that do not reference their objects. These verbs are mostly verbs of motion and perception (Jacques 2010).[8] An example is in (7) where the verb, *aro* 'have', only indexes the subject. The appearance of *-nɯ* referencing the plural object is ungrammatical. Additionally, both arguments of semi-transitives can be absolutive; the subjects of these non-agreeing verbs need not be flagged with the ergative postposition.[9] This is clear from (8) where *p^hama* 'parents' is the absolutive

[8] According to Dahlstrom (2013), the pseudotransitive verbs in Algic are also low in transitivity expressing possession, location, etc.

[9] More rarely. the ergative postposition is used, at least with some semi-transitive verbs (Jacques 2019a).

subject of the semi-transitive verb, *βgoz* ‘organise’ (translated by a passive to capture the information structural import of the fronted object).

JAPHUG

(7) Aʑo tɤ-rɟit χsɯm
1SG INDEF.POSS-child three
aro-a/*aro-a-nɯ.
have:FACT-1SG/have:FACT-1SG-PL
‘I have three children.’ (Jacques 2016:3)

(8) Ndʑi-stɯmmɯ nɯ pʰama
3DU.POSS-marriage DEM parents
pɯ-βgoz pɯ-ŋu.
PFV-organise PST.IPFV-be
‘Their marriage was arranged by their parents.’ (Jacques 2019:131)

Clearly, the absence of agreement is not a quirk of the verbal morphology since case morphology is also affected. The object itself is responsible and is causing both the absence of agreement and absence of ergative marking on the subject. Following Dahlstrom (2009), I suggest that this is because the unreferenced object of the semi-transitive verb is OBJ_θ while regular transitive verbs take OBJ as in (9).

(9) Agreeing: < SUBJ, OBJ >
Non-agreeing: < SUBJ, OBJ_θ >

OBJ_θ is the function held by themes of secundative verbs in Japhug, which are also unindexed by the verb. One secundative ditransitive in Japhug is *mbi* ‘give’ (the language also has indirective ditransitives, Jacques 2012). It heads clauses that resemble monotransitives: the subject is ergatively marked as in (10). The morphology is the same and agrees with only the subject and R, as in (11). T is unmarked and unreferenced.

JAPHUG

(10) …tɯmɯkɯmpɕi kɯ pɯ́-wɣ-nɯ-mbi-a
heavens ERG PFV-INV-AUTO-give-1SG
ɕti.
be.ASSERTIVE:FACT
‘…heavens have given (it) to me.’ (Grossman et al. 2018:12)

(11) Ki ɲɯ-ta-mbi.
DEM IPFV-1>2-give
'I give this to you.' (Jacques 2012)

Moreover, antipasssivation of these ditransitives creates a clause that resembles a semi-transitive.[10] Under antipasssivation, R is suppressed leaving only the subject and T, OBJ_{θ}. The resulting clause is effectively semi-transitive: both arguments bear absolutive case and the verb only indexes the subject (Jacques 2014). This is clear from the examples in (12) and (13) below: both with the verb, *mbi* 'give'. T is undexed in (12) and, more significantly in (13), the subject lacks ergative marking.

JAPHUG
(12) Stoʁ nɯ-rɤ-mbi-a.
bean PFV-ANTIPASS-give-1SG
'I gave beans (to someone).' (Jacques 2014:23)

(13) ɯʑo nɯ-rɤ-mbi.
3SG AOR-ANTIPASS-give.
'S/he gave it away (to people).' (G. Jacques p. c.)

The unindexed object bears a different syntactic function, OBJ_{θ}, and this is why it is unreferenced in Japhug. The case morphology of Japhug makes overt what Dahlstrom (2009) theorised was covert in the syntax of pseudotransitive objects in the Algic languages. However, as the next section will demonstrate, there are other instances of sporadic object agreement to which this analysis cannot be applied. Instead, morphology is the only difference.

Agreeing	Non-agreeing
Transitive < SUBJ, OBJ > ERG ABS exx. (3), (5), (6)	Semi-transitive < SUBJ, OBJ_{θ} > ABS ABS exx. (7), (8)

Table 1: Transitive and semi-transitive subcategorization frames and examples

[10] However, this is not the whole story since the antipassives of these verbs may also behave as more like transitives with ergatively marked subjects and indexed themes though without full transitive morphology in what is an unusual and unique pattern in Japhug (Jacques 2019a).

3 Morphology is only difference

Sporadic agreement with objects is also found in the Papuan family, Timor-Alor-Pantar (TAP).[11] Taking Kui to represent TAP, the OBJ_θ analysis does not look to be possible. Instead, the unreferenced objects are full objects. This leaves morphology as the only difference. This was previously implied to be the case for sporadic absolutive agreement in the Nakh-Daghestanian language, Archi (Sadler 2016).

For Kui, the OBJ_θ analysis, at first blush, has every hope of validity. Certain monotransitive verbs do not index their objects; this is also the case for ditransitive themes. This could be because they both hold the grammatical function, OBJ_θ. However, this is not the case as will be explained below.

In transitive clauses, Kui only has agreement for objects and then only on some verbs (a little over half of the transitive verbs observed. There does not appear to be a semantic or any other basis to the two classes; the two are simply lexical, Windschuttel 2019a). The following examples show this, a non-agreeing verb in (14) and an agreeing verb in (15). In the agreeing class, there are two series of agreement prefixes with each verb root choosing one or the other, the example here coming from the more common patientive series.

	KUI				
(14)	Anin	dona	ool	blēs.	
	person	yesterday	child	hit	
	‘Someone hit the child yesterday.’				

(15)	Na	ool	ga-wel.
	1SG.SBJ	child	3.PAT-wash
	‘I bathe a child.’		

These same prefixes also index the subject on a very small number of intransitive verbs; one is below in (16) (see Windschuttel & Shiohara 2017 and Windschuttel 2019). Agreement in Kui does not appear to be pronominal, at least, in intransitive clauses, since it can co-occur with free pronouns as in (17). In any case, zero anaphora is common and NPs, whether indexed on the verb or not, are readily elided.

[11] This looks to be connected to the similar pattern in TNG, briefly mentioned in the introduction, possibly because it is inherited from a common ancestor (Windschuttel 2019a).

KUI

(16) Cucu ga-rik-i.
Cucu 3.PAT-sick-PFV
'Cucu was sick.'

(17) Aninnok aban mi-a,
people village IN-IPFV
na gap n-awar nanga.
1SG.SBJ PART 1SG.PAT-return NEG
'If there were people in Lerabaing, I wouldn't have come back.'

The attraction of the OBJ_θ analysis is the same: ditransitive themes are unindexed in Kui as well. Nonetheless, objects of non-agreeing verbs do not bear this grammatical function. While Kui does not have case morphology (apart from on pronouns) or productive diathetic processes, it is a configurational language with a verb phrase (VP) which defines grammatical functions. Ditransitive themes are not in the VP while all monotransitive objects are, representing different functions.

The different phrase structure rules for the c-structure in Kui and how they define grammatical functions are given below in (18) (I is negative *nanga*, and various TAM clitics like *lei* 'PFV'). These will be justified next.

(18)

IP	→	DP (↑SUBJ)=↓	I' ↑=↓	
I'	→	DP (↑OBJ_θ)=↓	VP ↑=↓	I ↑=↓
VP	→	DP (↑OBJ)=↓	V ↑=↓	
VP/I'/IP	→	AdvP ↓∈(↑ADJ)	VP/I'/IP ↑=↓	

The VP can be defined by the placement of the first part of the negative *gap* and certain other adverbs.[12] They must precede the VP (or, alternatively, some other phrasal category like I' or IP).[13] This is shown for *gap* in (19).

[12] Outside of this function, *gap* has a partitive meaning, 'one of' (see also Windschuttel 2019:§6.2.2.1).
[13] It is possible that this phrasal category may be generalised to XP and semantics prevent its adjoining to DPs, etc. There is also a different class of postposed adverbs. For further details, see Windschuttel (2019:§4.6).

Monotransitive objects are in the VP defined thus. This includes the objects of non-agreeing verbs as in this example.[14]

KUI

(19)	(Gap)	anin	(gap)	[ol	(*gap)	blēs]$_{VP}$	nanga.
	PART	person	PART	child	PART	hit	NEG

‘Someone didn’t hit the child.’ (elicited)

According to object agreement and the VP, R patterns with indexed monotransitive objects. The object prefix on *-ei* ‘give’ references R as in (20) just like the monotransitive object of the agreeing verb, *-baran* ‘kill’, in (21) (see Windschuttel 2019:§6.2.3). R is part of the VP as in (22) appearing after the adverb, *awoi* ‘again’, just as P does in (19). This points to them holding the same function, OBJ.

KUI

(20)	Memang	anin	doi	in-ei…
	indeed	person	money	1PL.EXCL.PAT-give

‘Indeed, they gave us money...’ (*doi* entry in Katubi et al. 2013)

(21)	Nyi-baran	nanga!
	1PL.EXCL.PAT-kill	NEG

‘Don’t kill us!’

(22)	Gai	ga-yool=mo	awoi		
	3	3-child=MED	again		
	[gala	ga-ya=mo		ga-gamir-i]$_{VP}$	lei.
	gala[15]	3-sister.in.law=MED		3.PAT-marry-PFV	PFV

‘He gave his child again to the woman in marriage.’ (Shiohara n.d.)

Ditransitive themes, on the other hand, are unreferenced by the verb as in (20). They are not part of the VP since they may be separated from it by adverbs, etc. This is shown by (22) and (23); the adverb, *awoi* ‘again’,

[14] Moreover, this is clearly not due to adjacency. For example, in (i), where the object is elided, gap appears felicitously next to the verb

(i)	Anin	gap	[blēs]$_{VP}$	nanga.
	person	PART	hit	NEG

‘The person didn’t hit (anybody).’

[15] *gala* is a difficult to analyse word. There is some reason for treating it as a preposition although this is typologically unusual for an SOV language (Windschuttel 2019:§6.2.3.5). In any case, it clearly does not form a constituent with T as shown by (22) and (24) and thus this should not directly affect its analysis.

intervenes between T and the VP in (22) and likewise for the first part of the negative *gap* in (23). Thus, T has its own GF distinct from R and P, OBJ_{θ}. This is not the function held by the objects of non-agreeing verbs. Unlike OBJ_{θ}, they fill the VP internal OBJ position as is clear from (24), repeated from (19) above. The unindexed object may not be followed by *gap* unlike OBJ_{θ}. This indicates that they are OBJ just like indexed objects.

	KUI					
(23)	(Gap)	na	(gap)	bat		(gap)
	PART	1SG.SBJ	PART	coconut		PART
	[gala	(*gap)	ø-ei]$_{VP}$		nanga.	
	gala	PART	2SG.PAT-give		NEG	
	'I didn't give you the coconut.' (elicited)					

(24)	(Gap)	anin	(gap)	[ol	(*gap)	blēs]$_{VP}$	nanga.
	PART	that	PART	child	PART	hit	NEG
	'Someone didn't hit the child.' (elicited)						

There is one other process that plausibly picks out OBJ_{θ} to the exclusion of other objects. NP-fronting is observed with all objects whether referenced or unreferenced, including recipients of ditransitive verbs, except never T, that is, OBJ_{θ} (just like OBJ_{θ} in English, the recipient in double verb constructions, see Huddleston 1984:195-203).

Monotransitive objects, whether indexed or unindexed, can be fronted as in (25) and (26). In ditransitive clauses, R can also be fronted as in (27) while T has not been seen fronted (the possibility must be considered, however, that this could be a gap in the data rather than a hard constraint). If this test is valid, objects of non-agreeing verbs pattern again with indexed P and not T/OBJ_{θ}.

	KUI				
(25)	[Na-gaj]$_{OBJ}$		anin	ga-marei	.
	1SG-wage		person	3.PAT-go.up	
	sampe	rib		asaga	yesanusa
	until	thousand		hundred	nine
	'My wage was raised to 900 000 rupiah.'				

(26) Na palak og el-i.
1SG.SBJ land PROX buy-PFV
[Palak og]$_{OBJ}$ na el lei…
land PROX 1SG.SBJ buy COMPL
'I bought this land. Having bought this land I…'

(27) [Palak=gog]$_{OBJ}$ na gala g-ei.
earth=TOP 1SG.SBJ *gala* 3.PAT-give
'I gave (money) to the earth lord.' (Katubi et al. nd)

Thus in Kui, indexed and unindexed objects are treated the same syntactically having the OBJ function, both part of the VP.[16] The only difference between the two is the morphology of the verb. Thus morphology driven sporadic agreement does appear to be a necessary category corresponding to a different set of grammars.

4 Morphological solutions

There could be a number of ways to represent the difference in morphology between agreeing and non-agreeing verbs in Kui. There are two main families of morphological theories, incremental and realisational, defined by Toivonen (2002) as follows: "in incremental theories, morphosyntactic information gets added incrementally as morphemes are added to a stem. In a realizational theory, a word's association with certain morphosyntactic properties licenses the appropriate affixes."

The incremental approach has a long history in the LFG (from as far back as Simpson 1983, for example, and assumed in Bresnan 2001). Individual morphemes are given lexical entries and are combined together by sublexical rules. Following Schwarze's (1999) approach to the inflectional classes in Italian, an f-structure CLASS feature could be created to represent the different verbal agreement classes in Kui. This along with some other features are given below in (28) for an agreement prefix and verb roots from two of the classes. The features of the prefix and root would give rise to the same f-structure according to the sublexical rule in (29). Coherence would forbid the prefix *na-* with the CLASS value PAT from being present in the same structure as the non-agreeing verb, *-tak* 'feed', which has a different

[16] Rachel Nordlinger suggested that relativisation could distinguish OBJ and OBJ_θ (according to the Accessibility Hierarchy, Keenan & Comrie 1977). Unfortunately, the data is not available for Kui, relative clauses being rare and where present typically created with a borrowed relativiser.

value for the feature, namely, ¬AGR (PAT represents the patientive prefix series).

(28) na- aff ($\uparrow$ PERS) = 1, ($\uparrow$ NUM) = SG, ($\uparrow$ CLASS) $=_c$ PAT

-as V_{root} ($\uparrow$ PRED) = ‘-as<SUBJ OBJ>’, ($\uparrow$ CLASS) = PAT

-tak V_{root} ($\uparrow$ PRED) =‘-tak<SUBJ OBJ>’, ($\uparrow$ CLASS) = ¬AGR

(29) V $\rightarrow$ $\underset{\uparrow=\downarrow}{\text{aff*}}$ $\underset{\uparrow=\downarrow}{V_{root}}$ $\underset{\uparrow=\downarrow}{\text{aff}}$

More recently, realisational theories have become popular in LFG circles (Sadler & Spencer 2001, Sadler & Nordlinger 2004, Dalrymple 2015, etc. though see Andrews 2019 for an exception). Following the thought of Windschuttel (2012:14), the non-agreeing verbs could be treated as “morphological intransitives”, that is, essentially intransitive deponents. Sadler and Spencer (2001) presented an analysis of Latin passive deponents in Paradigm Function Morphology (Stump 2001, 2016). They proposed a rule of referral generating semantically active meanings from morphologically passive forms (from the s-paradigm to the m-paradigm). In (30) their notation is reformulated to express a referral from transitive to intransitive for the non-agreeing verbs (see Brown 2015 for a prior decomposition of ‘transitivity’ into s-features and m-features for similar purposes).

(30) (s-Transitivity:TRANS) $\Rightarrow$ (m-Transitivity:TRANS)

$\Downarrow$

(m-Transitivity:INTRANS)

This analysis has the added advantage that it could be reversed to account for agreement prefixes on intransitive verbs. They could be regarded as intransitive verbs with transitive morphology (somehow the grammatical function indexed would need to be changed to the only argument, SUBJ; perhaps an OT analysis similar to Alsina & Vigo 2017 could be developed with constraints, AGROBJ, for agreement with objects, and AGRSHARE, for agreement with any argument. These would be ordered AGROBJ >> AGRSHARE so that prefixal agreement would preferentially index an object but the presence of AGRSHARE would ensure that an argument is indexed where no object is present).[17]

Whatever the details of a morphological analysis, it is clear that one is necessary and an analysis based on grammatical relations like Section 2 is not possible for the non-agreeing verbs in Kui.

5 Conclusion

There are thus two types of sporadic object agreement. In one type, the agreeing and non-agreeing verbs subcategorise for two different types of objects, OBJ and OBJ_θ following Dahlstrom (2009); this difference in syntax affects more than just agreement. In the other type, only morphology distinguishes the verbs that agree and do not agree. In particular, the OBJ vs.. OBJ_θ analysis is not valid in Kui where both are OBJ and morphology is the difference. These look to represent real typological differences.

The difference between morphological and syntactic sporadic agreement parallels the distinction between syntactic and morphological ergativity and their analyses in LFG. Syntactic ergativity was contrasted with accusativity by how SUBJ and OBJ were linked by Manning (1996) and Arka & Manning (1998). The correspondences between thematic roles and grammatical functions were reversed as in (31). This is distinguished from morphological ergativity which is not based on basic grammatical functions. Instead, it may be directly stipulated with the equations in the f-structure of transitive verbs in (32) (Sadler 2016).[18] Both syntactic ergativity and sporadic agreement are analysed through regular grammatical functions while the morphological analogues require other solutions in both cases.

The difference between morphological and syntactic sporadic agreement parallels the distinction between syntactic and morphological ergativity and their analyses in LFG. Syntactic ergativity was contrasted with accusativity

[17] Nordlinger (2010) and Windschuttel (2012) did provide an analysis of similar impersonal constructions where ‘object’ agreement indexes the subject. They suggested that ‘object’ agreement is actually ambiguous between SUBJ and OBJ, the object reading being forced in transitive clauses by the co-occurring subject agreement. However, their analysis could not be carried over to Kui since there is no subject agreement in transitive clauses.

[18] Falk (2006) treats ergativity quite differently, however, and while he does not spell out the analysis of morphological ergativity in detail, it is also not based directly on grammatical functions unlike his analysis of syntactic ergativity. Sadler (2016) in addition to the analysis of (32) also proposes an alternative analysis of Archi using the PIV function, which, if used in Falk’s (2006) original sense, would imply that the ergativity is syntactic.

by how SUBJ and OBJ were linked by Manning (1996) and Arka & Manning (1998). The correspondences between thematic roles and grammatical functions were reversed as in (31). This is distinguished from morphological ergativity which is not based on basic grammatical functions. Instead, it may be directly stipulated with the equations in the f-structure of transitive verbs like in (32) (Sadler 2016).[19] Both syntactic ergativity and sporadic agreement are analysed through regular grammatical functions while the morphological analogues require other solutions in both cases.

(31) ERG: < OBJ, SUBJ >
ag pt
ACC: < SUBJ, OBJ >

(32) (↑SUBJ CASE) = ERG, (↑OBJ CASE) = ABS

An interesting area for further research on sporadic verb agreement would be sign languages. This is arguably the most notable instance of sporadic verb agreement since it is found in all sign languages with agreement (Mathur & Rathmann 2012, Aronoff et al. 2005, Steinbach 2011). Agreement in sign languages is by physically 'indexing' the arguments, that is, pointing at the referent in real space. Figure 1 displays an example of this in BSL (British Sign Language). While this has long been analysed as agreement (from as early as Meier 1983, Padden 1988), this analysis has recently been challenged (from Liddell 2000 to Schembri 2018), its sporadic nature being only one of many unusual features (Aronoff et al. 2005, Lillo-Martin & Meier 2011). However, should the traditional analysis prove correct, it will be interesting to see how the non-agreeing verbs in these languages fit into the typology introduced in this paper.

[19] Falk (2006) treats ergativity quite differently, however, while the analysis of morphological ergativity is not spelt out in detail but it is also not based directly on grammatical functions in the same way syntactic is. Sadler (2016) also proposes an alternative analysis of Archi using the PIV function, which, if used in Falk's (2006) original sense, would imply that the ergativity is syntactic.

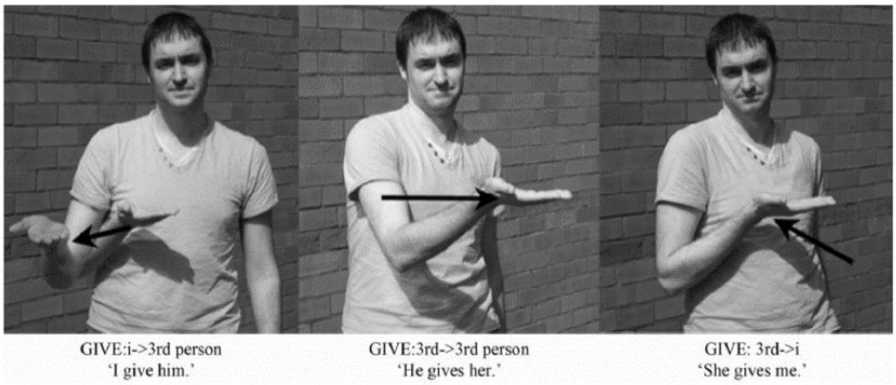


Figure 1: The agreeing verb GIVE in BSL. The direction of the arrow shows the direction of motion, from subject to object in this case (Nick Palfreyman in De Vos 2012:122-3)

Casting the net still wider, these instances of sporadic agreement have been linked to other constructions: transitivity discord constructions (Zúñiga 2019) and differential object marking (Klamer & Kratochvíl 2018). These are, in my view, significantly different since they meaningfully alternate with the same root rather than lexically classifying different verb roots. Despite this difference, it may be possible to develop a similar morphology vs. syntax typology for these phenomena as well, with similar implications for their analyses in LFG (Çetinoğlu & Butt 2008 is already an example of the syntactic analysis).

References

Alsina, Alex and Eugenio M. Vigo. 2017. Agreement: interactions with case and raising. In M. Butt and T. King (eds.) *Proceedings of the LFG17 Conference*, 3-23. Stanford, CA: CSLI.

Andrews, Avery. 2019. Incremental Morphology in LFG. Presentation given at LFG19 Conference, June 2019.

Arka, I. Wayan and Christopher D. Manning. 1998. Voice and grammatical relations in Indonesian: A new perspective. In M. Butt and T. King (eds.) *Proceedings of the LFG98 Conference*, 1-17. Stanford, CA: CSLI.

Aronoff, Mark, Irit Meir and Wendy Sandler. 2005. The paradox of sign language morphology. *Language* 81(2):301-344.

Belyaev, Oleg. 2013. Optimal agreement at m-structure: Person in Dargwa. In M. Butt and T. King (eds.) *Proceedings of the LFG13 Conference*, 90–110. Stanford, CA: CSLI.

Brown, Dunstan. 2015. An implemented account of the antipassive / transitive mismatch in Chukchi. Ms. The University of York.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Çetinoğlu, Özlem and Miriam Butt. 2008. Turkish non-canonical objects. In M. Butt and T. King (eds.) *Proceedings of the LFG08 Conference*, 214-234. Stanford, CA: CSLI.
Corbett, Greville G. 2006. *Agreement*. Cambridge: CUP.
Dahlstrom, Amy. 2009. OBJ_θ without OBJ: A typology of Meskwaki objects. In M. Butt and T. King (eds.) *Proceedings of the LFG09 Conference*, 222–239. Stanford, CA: CSLI.
Dahlstrom, Amy. 2013. Argument structure of quirky Algonquian verbs. In T. King and V. de Paiva (eds.) *From quirky case to representing space: Papers in honor of Annie Zaenen*, 61–71. Stanford, CA: CSLI.
Dalrymple, Mary. 2015. Morphology in the LFG architecture. In M. Butt and T. King (eds.) *Proceedings of the LFG15 conference*, 64–83. Stanford, CA: CSLI Publications.
De Vos, Connie. 2012. Sign-Spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. Nijmegen: PhD Dissertation at Radboud University.
Falk, Yehuda. 2006. Subjects and Universal Grammar: An explanatory theory. Cambridge: CUP.
Fedden, O. Sebastian. 2017. Niches and systems of sporadic agreement, 48th Annual Meeting of the Societas Linguistica Europaea (SLE), University of Zurich, September 2017.
Fedden, O. Sebastian. 2017a. Überlegungen zu sporadischer Kongruenz, Vielfaltslinguistik-Konferenz, Universität Leipzig, March 2017.
Fedden, O. Sebastian. 2017b. Steps towards a typology of sporadic agreement systems, Typologie et modélisation des systèmes morphologiques, University of Chicago Center in Paris, January 2017.
Fedden, O. Sebastian. 2019. To Agree Or Not To Agree? - A Typology of Sporadic Agreement. In M. Baerman, O. Bond and A. Hippisley (eds.) *Morphological Perspectives: Papers in Honour of Greville G. Corbett*. Edinburgh: Edinburgh University Press.
Foley, William. 1986. *The Papuan Languages of New Guinea*. Cambridge: CUP.
Foley, William. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29:357-404.
Grossman, Eitan, Anton Antonov and Guillaume Jacques. 2018. A cross-linguistic rarity in synchrony and diachrony. *Language Typology and Universals* 71(4):513-538.
Huddleston, Rodney 1984. *Introduction to the Grammar of English*. Cambridge: CUP.
Jacques, Guillaume. 2004. Phonologie et morphologie du japhug (Rgyalrong). PhD dissertation, Université Paris VII – Denis Diderot.
Jacques, Guillaume. 2010. The inverse in Japhug Rgyalrong. *Language and Linguistics* 11(1):127-157
Jacques, Guillaume. 2012. Argument demotion in Japhug Rgyalrong. In Katharina Haude and Gilles Authier (eds.) *Ergativity, Valency and Voice*, 199-226. Berlin: Mouton De Gruyter.
Jacques, Guillaume. 2014. Denominal affixes as sources of antipassive markers in Japhug Rgyalrong. *Lingua* 138:1-22.

Jacques, Guillaume. 2014a. Clause linking in Japhug Rgyalrong. *Linguistics of the. Tibeto-Burman Area* 37(2):263-327.
Jacques, Guillaume. 2016. Subjects, objects and relativization in Japhug. *Journal of Chinese Linguistics* 44(1):1-28.
Jacques, Guillaume. 2019. Verbal valency and Japhug / Tibetan language contact. *Journal of Language Contact* 12:116-140.
Jacques, Guillaume. 2019a. A grammar of Japhug (verbal morphology). Ms. https://www.academia.edu/38793596/A_grammar_of_Japhug_verbal_morphology_
Katubi, Thung Ju Lan and Marcelinus Yeri Fernandez Akoli. nd. Kui corpus. Indonesian Institute of Science (LIPI).
Katubi, Thung Ju Lan and Marcelinus Yeri Fernandez Akoli. 2013. *Kamus kecil bahasa Kui* [A small dictionary of the Kui language]. Jakarta: LIPI Press
Keenan, Edward L. and Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63-99
Klamer, Marian and František Kratochvíl. 2018, The evolution of differential object marking in Alor-Pantar languages. In I. A. Seržant and A. Witzlack-Makarevich (eds.) *Diachrony of differential argument marking*, 69-95. Berlin: Language Science Press.
Liddell, Scott K. 2000. Indicating verbs and pronouns: Pointing away from agreement. In E. Karen and L. Harlan (eds.) *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 303–320. Mahwah, NJ: Lawrence Erlbaum Associates
Lillo-Martin, Diane and Richard P. Meier. 2011. On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics* 37(3/4):95-141.
Manning, Christopher D. 1996. *Ergativity: Argument Structure and Grammatical Relations*. Stanford: CSLI.
Mathur, Gaurav and Christian Rathmann. 2012. Verb agreement. In R. Pfau, M. Steinbach and B. Woll (eds.) *Sign language — An international handbook*, 136-157. Berlin: De Gruyter Mouton.
Meier, Richard P. 1982. Icons, analogues, and morphemes: The acquisition of verb agreement in ASL. San Diego: PhD Dissertation at University of California.
Nordlinger, Rachel. 2010. Agreement Mismatches in Murrinh-Patha Serial Verbs. In Yvonne Treis and Rik De Busser (eds.), *Selected Papers from the 2009 Conference of the Australian Linguistic Society*.
Padden, Carol. 1988. *Interaction of Morphology and Syntax in American Sign Language*. New York: Garland.
Pilhofer, Georg. 1933. *Grammatik der Kâte-Sprache in Neuguinea*. Vierzehntes Beiheft zur Zeitschrift für Eingeborenen-Sprachen. Berlin: Dietrich Reimer.
Rhodes, Richard. 1990. Secondary objects in Ojibwe. In K. Dziwirek, P. Farrell, and E. Mejías-Bikandi (eds.) *Grammatical Relations: A Crosstheoretical Perspective*, 401–414. Stanford, CA: CSLI.
Sadler, Louisa and Rachel Nordlinger. 2006. Case stacking in realizational morphology. *Linguistics* 44(3):459-487.
Sadler, Louisa, and Andrew Spencer. 2001. Syntax as an exponent of morphological features. In G. Booij and J. van Marle (eds.) *Yearbook of Morphology 2000*, 71-96. Dordrecht: Kluwer.

Sadler, Louisa. 2016. Agreement in Archi: An LFG perspective. In O. Bond, G. Corbett, M. Chumakina and D. Brown (eds.) *Archi: Complexities of agreement in cross-theoretical perspective*, 150-183. Oxford: OUP.

Schembri, Adam, Kearsy Cormier and Jordan Fenlon. 2018. Indicating verbs as typologically unique constructions: Reconsidering verb 'agreement' in sign languages. *Glossa: A Journal of General Linguistics* 3(1):89.

Schwarze, Christoph. 1999. Lexical-functional morphology and the structure of the lexicon. In L. Mereu (ed.) *Boundaries of Morphology and Syntax*, 73-95. Amsterdam: John Benjamins.

Shiohara, Asako. 2010. Verb Types and Semantic Roles of Person Markers of Kui, Alor. Paper given at Workshop on the Languages of Papua 2, 8-12 February.

Shiohara, Asako. n. d. Transcription of story entitled Dandagesin.

Simpson, Jane H. 1983. Aspects of Warlpiri Morphology and Syntax. PhD thesis, Massachusetts Institute of Technology.

Spencer, Andrew. 2009. Realization-based morphosyntax: the German genitive. In P. Steinkrüger and M. Krifka (eds.) *On Inflection. Trends in Linguistics, Studies and Monographs 184*, 173-218. Berlin: Mouton de Gruyter.

Steinbach, Markus. 2011. What Do Agreement Auxiliaries Reveal About the Grammar of Sign Language Agreement? *Theoretical Linguistics* 37:209-221.

Stump, Gregory. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: CUP.

Stump, Gregory. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface.* Cambridge: CUP.

Suter, Edgar. 2012. Verbs with pronominal object prefixes in Finisterre-Huon languages. *Language and Linguistics in Melanesia*. Special Issue(2012):23–59.

Toivonen, Ida. 2002. Review of Stump, Inflectional Morphology. https://linguistlist.org/issues/13/13-622.html

Vincent, Alex. 2003. Dialects of the Tairora Language of the Eastern Highlands of Papua New Guinea. Manuscript.

Windschuttel, Glenn and Asako Shiohara. 2017. Kui. In A. Schapper (ed.) *Papuan languages of Timor-Alor-Pantar: Sketch grammars Volume 2*, 109-183. Berlin: Mouton de Gruyter.

Windschuttel, Glenn. 2012. The impersonals of Yagaria. Honours thesis: University of Sydney

Windschuttel, Glenn. 2017. Object prefixing verbal subclass as a diagnostic of TNG affiliation. Paper given at Workshop on the Languages of Papua 4, 23-26 January 2017.

Windschuttel, Glenn. 2018. A typology of uninflectedness. Manuscript, University of Newcastle.

Windschuttel, Glenn. 2019. Extended Sketch of Kui. Manuscript, University of Newcastle.

Windschuttel, Glenn. 2019a. Object verbs in Kui, TAP and TNG. PhD thesis, University of Newcastle, Australia.

Zúñiga, Fernando. 2019. *Grammatical Voice*. Cambridge: CUP.

Aspectual Reasoning in LFG – A Computational Approach to Grammatical and Lexical Aspect

Mark-Matthias Zymla
University of Konstanz

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 353–373

http://csli-publications.stanford.edu/LFG/2019

Keywords: tense/aspect, ParTMA, annotation, XLE

Zymla, Mark-Matthias. 2019. Aspectual Reasoning in LFG – A Computational Approach to Grammatical and Lexical Aspect. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 353–373. Stanford, CA: CSLI Publications.

Abstract

This paper provides a semantic analysis of grammatical aspect built on the foundation of Lexical Functional Grammar. Due to this dependency this paper will also pay attention to the morphosyntactic analysis of aspect in LFG. The main focus lies on the treatment of aspect in the computational LFG grammars produced within the Xerox Linguistic Environment. The results of this endeavour are two-fold: firstly, an evaluation of the capabilities of the English XLE grammar with respect to aspect as a grammatical feature. Secondly, a semantic theory of grammatical aspect that is couched within the broader ParTMA effort with the goal of providing a cross-linguistically viable annotation and representation of tense and aspect.

1 Introduction

The main goal of this paper is to derive a semantic analysis of grammatical aspect from LFG's c- and f-structure. The underlying syntactic representation is that provided by the English LFG grammar written in the Xerox Linguistics Environment (XLE). The semantic representation is derived via annotation rules as described in Zymla (2017a,b), which are inspired by the idea of packed rewriting, the underlying concept of the XLE transfer system (Crouch et al. 2017). This annotation is a part of the ParTMA project, which is a daughter project of the ParGram effort (Butt et al. 2002). The English XLE grammar is maintained as part of this project.

Although both tense and aspect have been acknowledged as an integral part of reasoning, corresponding efforts within the domain of computational grammar development and computational linguistics in general are rather meager. Building on the ParTMA project, I present a semantic extension of the XLE analysis of grammatical aspect. This extension serves as a foundation for a semantic representation enriched with aspect information. I set a modest benchmark for this kind of representation, namely that of being able to deal with the inference patterns underlying the imperfective paradox (Dowty 1977). This pattern is illustrated in (1). The progressive as an instance of imperfective aspect allows the inference of its perfective counterpart in some cases but not in others.

(1) a. John was drawing a circle $\not\to$ John drew a circle.
b. John was pushing a cart $\to$ John pushed a cart.

It is undeniable that there is a vast amount of literature on this topic, however, concrete implementations are lacking. In this paper, I provide an overview of the syntactic and semantic category *imperfective* and related properties and show how a computationally viable semantic representation can be acquired that allows for

[†]I thank the Nuance foundation who funded the project *Tense and Aspect in Multilingual Semantic Construction* as well as the project *VALIDA* for supporting my research. Furthermore, I thank Miriam Butt, Maribel Romero and Aikaterini-Lida Kalouli for helpful discussion. Finally, I thank the participants of the LFG2019 conference as well as the reviewers for comments and suggestions.

automated reasoning and that captures the inference pattern illustrated above. Concretely, the semantic representation presented here is an extension of the computational semantic formalism for deriving abstract knowledge representations (AKR) (Bobrow et al. 2007).

The paper is structured as follows: In the first part of the next section, I discuss the treatment of grammatical aspect in the ParGram grammars (Butt et al. 2002) and the English grammar in particular. In the second part of Section 2, I enrich (an idealized form of) the English grammar with aspect information in the spirit of the ParTMA annotation scheme. In Section 3, I provide an overview of the relevant semantic properties from a formal perpsective, and Section 4 fleshes out these properties in the computationally viable formalism of AKR. Section 4 also explains how the new extension of AKR covers the reasoning patterns illustrated above. Section 5 concludes.

2 Viewpoint Apsect at the Syntax/Semantics Interface

In this paper, I adopt the (very general) cross-linguistic picture for grammatical aspect features shown in Figure 1 (see Comrie (1976) for reference; Carlson (2012) for comments). I focus on imperfective aspect as a central topic of this paper.

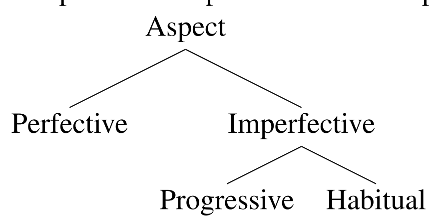


Figure 1: Aspect across languages

2.1 Morphosyntactic Realization of Imperfective Aspect in English

In languages, where there is no overt habitual/progressive distinction the imperfective may be used to express both. This is for example the case in Greek and Italian (Ferreira 2016).

Languages that do not subsume a progressive interpretation and a habitual interpretation under the banner of the imperfective often use the grammatically least marked verb form to express habituals (Dahl 1985, Carlson 2012). This is for example the case in English, where habituality is (although not exclusively) associated with present tense morphology as illustrated in (2-b).

(2) a. John is smoking a cigarette. *progressive*
    b. John smokes cigarettes. *habitual*

While the progressive in (2-a) refers to an episodic or incidental interpretation (Carlson 2012), the habitual sentence expresses that John regularly smokes cigarettes. However, it does not commit to the fact that John is smoking at the time of speaking, while the progressive sentence does. This consequently means that the *progressive* and the *habitual* may be expressed in a morphosyntactically distinct manner.[1]

In the English ParGram grammar the Progressive is treated as a functional attribute. Thereby, the Auxiliary takes the role of a feature carrying element. By projecting morphological features to a morphological structure, certain dependencies are ensured, e.g., perfect vs. progressive auxiliary (Butt et al. 1999, 1996). Thus, the progressive and the perfect are analyzed as a binary feature:

(3) $$\left[\text{TNS-ASP} \quad \left[\text{PROG +\_/-\_ , PERF +\_/-\_}\right]\right]$$

As pointed out above, the *habitual* is associated with the simple present in English. This would correspond to the following collection of `TNS-ASP` features:

(4) $$\left[\text{TNS-ASP} \quad \left[\text{TENSE pres, PROG -\_ , PERF -\_}\right]\right]$$

However, this configuration does not automatically warrant a habitual interpretation. Thus, the habitual is correctly not treated as a functional attribute in the English ParGram grammar. In Section 3, I will capture the intuition that the habitual is a semantic category rather than a morphosyntactic one.

2.2 A Brief Note on Perfective and Perfect Forms in English

For the sake of this paper, I make explicit the assumption from the discussion of example (1) that the simple past in English has a default perfective interpretation. In the case of English this elicits an interpretation that entails a boundary point or point of completion. This is not to be confused with the perfect which is introduced by an auxiliary carrying tense information in combination with a past participle form of the main verb. In many languages similar constructions have a past tense interpretation in addition to some aspectual interpretation. This is for example the case in German where the perfect has been argued to be ambiguous (Löbner 2002, Musan 2002). The perfect in English is often treated as an aspectual category on the basis that it can be combined with the progressive. This is in line with the observation that aspect is a recursive category (De Swart 1998, de Swart 2016).

In a ParTMA style annotation scheme, tense and aspect information are disentangled allowing us to annotate fine-grained differences between languages with respect to such categories. However, for reasons of space I will not discuss the distinctive annotation of perfect vs. perfective in this paper.

[1] As some researchers point out, the progressive form in English can also mark a habitual sentence. The example below by Steedman (1997) illustrates this. The quantificational adverbial phrase "these days" enforces a habitual interpretation (section 4).

(i) I am writing a sonnet in fifteen minutes (these days).

2.3 Semantic labeling

This paper employs the strategy introduced and refined in Zymla (2017a,b) to map syntactic information onto semantic information. For our current purposes, the rules remain fairly straightforward. The labels `#g, #h, ...` refer to f-structure nodes, while `#g', #h', ...` refer to a semantic projection of the respective f-structure. This is not the typical s-structure assumed for Glue semantics (Dalrymple 2001), but rather a specific tense and aspect meaning structure.[2] A basic example for the domain of tense is shown in (5).

(5) `#g TNS-ASP #h TENSE past` $\rightarrow$ `#g' TEMP-REF` $t \prec t_0$

The rule simply states that past tense morphology induces a past tense interpretation. Zymla (2017b) illustrates how to resolve situations where this rule is bound to fail, namely additional rules can rewrite basic interpretations in specific contexts, resulting in a layered annotation.

In the next section, I start with a subset of rules for TA in English. The rules are illustrated below. First, a progressive label simply introduces an ongoing interpretation. Then, as already pointed out, (one) expression of habituality in English is associated with the present tense. However, this only works with *events* (see Section 3) and some information seems to still be missing. I am not committing to any particular approach at this point but see Ferreira (2016) for semantic sg/pl distinction of the verb as a pontential candidate.[3] The missing element is illustrated as "..." in the rules below.

(6) Imperfective categories (exemplified)

```
#g TNS-ASP #h PROG +_ → #g' ASPECT impv,
                        #g' ASPECT-RESTR ongoing

#g TNS-ASP #h TENSE pres,
#h PROG -_,#h PERF -_,
#g' ASP-CLASS event,...    → #g' ASPECT impv,
                             #g' ASPECT-RESTR hab
```

With respect to perfective constructions, I only make use of the simple past with a perfective interpretation. For this purpose, the following rule suffices for the understanding of this paper.

(7) Perfective category (exemplified)

a.
```
#g TNS-ASP #h TENSE past,
#h PROG -_,#h PERF -_
                      → #g' ASPECT prv,
                      #g' ASPECT-RESTR unspec
```

[2]Reconciling the two structures in a computational Glue approach is left for future work since it is orthogonal to the present paper

[3]See also Bertinetto & Lenci (2012) for a more general overview.

The choice for the semantic labels is partly inspired by the structure in Figure 1. However, it is primarily affected by the specific choice of semantic analysis for viewpoint aspect (section 3). This means that, in the next section, the exact semantic interpretation of the labels above is illustrated. As is typical for annotation in terms of semantic labels, multiple semantic theories can make appropriate use of the given annotation to generate semantic representations (Ide & Pustejovsky 2017). In this paper, I opt for a computational semantic formalism developed specifically for LFG, namely Abstract Knowledge Represenations (AKR) (Bobrow et al. 2007). However, I motivate the computational analysis in terms of formal semantic insights. These take the center stage in the next section.

3 Interpreting the Semantic Annotation

In the previous section I provided a set of semantic labels which are derived from specific syntactic configurations. In this section, I will discuss a semantic composition process that incorporates the semantic aspect annotation.

So far, I have contrasted imperfective syntactic and semantic categories with perfective categories. As pointed out in the introduction, the perfective is generally associated with linguistic expressions that describe a completed event. Completion is a property of *telic* predicates – the property of having an inherent endpoint. In this paper I take the stance that telicity is a conceptual property that is independent from grammatical aspect, which is associated with temporal boundaries. The following examples (8) and (9) by Depraetere (1995) summarize this. They are to be read as follows: If a predicate with an inherent endpoint reaches its endpoint, it is automatically temporally bounded by that endpoint. If the endpoint is not reached, then there is no temporal boundary. On the other hand, expressions with no inherent endpoint (e.g., states such as *John loves Mary* or processes such as *John is drinking beer*) can either be temporally bounded or unbounded.

(8) +inherent/intended endpoint

a. + endpoint reached: + temporal boundary
b. - endpoint reached: - temporal boundary

(9) - inherent/intended endpoint

a. + temporal boundary
b. - temporal boundary

I assume that the perfective is a boundary operator in the spirit of Depraetere (1995). However, what does it apply to? Let us start with a three-way distinction for situations: states, processes, and events (de Swart 2016, Filip 2012).

States and processes have no inherent intended enpoint, while events do so. This means applying the perfective to an event results in an event description of which the event's result is a part of, according to Depraetere (1995). A popular set of theories following Reichenbach (1947) analyses grammatical aspect as a

		durative	change	endpoint
	state	+	-	-
process	activity	+	+	-
event	achievement	-	+	+
	accomplishment	+	+	+

Table 1: Vendler (1957) classification from Filip (2012)

relation between the reference time, i.e., the time a linguistic expression refers to and the run-time of the event or eventuality time. Steedman (1997) points out that many theories fall into a similar schema, which I also follow. A refinement of Reichenbach's (1947) analysis provided by Klein (1994) realizes this relation as a subset relation. In the following example, *R* denotes the reference time and $\tau(e)$ refers to the run time of the given event.

(10) PERFECTIVE: $\tau(e) \subseteq R$
PROGRESSIVE: $\tau(e) \supset R$

This analysis causes some issues both for the perfective and the progressive. However, as their core insights remain valid, they provide an appropriate basis for this paper. However, for the imperfective, I provide a more refined analysis based on recent advances in the formal literature to support the computational implementation proposed in this paper.

3.1 Semantic Analysis of the Progressive

Formal semantic research of the imperfective has a long-standing tradition traced back to Dowty (1977). As pointed out in the introduction, the challenge provided by the progressive form in English is encapsulated in the inference pattern in (11), repeated below for convenience:

(11) a. John was drawing a circle $\not\rightarrow$ John drew a circle.
b. John was pushing a cart $\rightarrow$ John pushed a cart.

Example (10) suggests that only part of the given event in (11-a) is described by the progressive form. The beginning and endpoint of the event in question are not *visible* (cf. Smith (1991)). However, this non-visibility is non-existence in some cases. Thus, in (12) we use an expression that implies the complete crossing of the road but this implication is defeasible, as shown by the second clause of the sentence (Asher 1992).

(12) The chicken was crossing the road, when it was hit by a truck.

Processes which do not have an inherent endpoint, as in (11-b), may be bounded (perfective) or unbounded (progressive), but their internal structure remains the same. Thus, the bounded interpretation could potentially be derived as a subpart of the progressive interpretation. This explains why the inference in (11-b) is valid.

There are two major camps when it comes to analyzing the progressive and the corresponding inferences with respect to events and processes. In the one camp, an analysis based on the part-whole structure of events as well as their NP arguments (e.g., Krifka (1998), Parsons (1990)) is proposed. In the other camp, people have provided an intensional analysis of the progressive (e.g., Dowty (1977), Landman (1992)). An extensive discussion of both camps is provided in Portner (2019).

In this paper, I opt for an analysis based on the ideas of the second camp. In particular, I treat the progressive as a quantifier over situations along the lines of Cipria & Roberts (2000) and Arregui et al. (2014), especially the latter. The reason for this is simple: It provides a unified analysis over different meanings encoded in imperfective categories that is straightforwardly implementable based on the semantic labels proposed above. Concretely, the `ASPECT` label with the value `impv` introduces a quantifier over situations:

(13) **impv:** $\lambda M.\lambda P.\lambda s.\forall s'[M(s)(s') \rightarrow \exists e[P(e)(s')]]$

M stands for an accessibility relation that functions as a modal base in terms of Kratzer et al. (1991). For the present paper, I have introduced the values `ongoing` and `hab`, which introduce an appropriate relation to saturate *M*. Arregui et al. (2014) propose, among others, the following modal bases:

(14) a. $MB_{ongoing} : \lambda s.\lambda s'.s' < s$ (s' is part of s)
b. $MB_{gen} : \lambda s.\lambda s'.$ s' is a characteristic part of s

The modal base in (14) corresponds roughly to the formalization in (10). (14-b) is provided for generic/habitual sentences with disregard of fine semantic differences (see Carlson (2012) for an overview). It is noteworthy, that s corresponds to the topic situation, i.e., the actual world. This means that, if the truth of $P(e)$ is given in s', $P(e)(s)$ is automatically true as well.

This only works under the assumption that we can concretely define characteristic parts of s. This is fairly difficult for a computational system since it involves both contextual and world knowledge. For this reason, I propose a variant of the modal base that is relative to P.[4] This accounts for certain fringe cases as well as unreliability effects (see Fara (2005)), but, more importantly, the failure of entailment between a present tense habitual sentence and its present progressive counterpart is straightforwardly captured. This topic is picked up again in Section 4.

(15) MB_{gen} given $P : \lambda s.\lambda s'.$ s' is a characteristic P situation, accessible from s.

Finally, (14-a) explains the progressive for processes, but it fails to capture the

[4]See Arregui et al. (2014) for a similar proposal for progressives.

defeasibility of the result state implication that is part of the meaning of the progressive when applied to events. For this purpose, the modal base in (16) was designed. The basic idea goes back to Dowty (1977). The takeaway that is relevant for this paper is simply that the progressive form expresses a part of a complete event, the completion of which might take place in a possible world rather than the actual world. However, the progressive form is derived from a hypothetical complete form of the corresponding eventuality.

(16) $MB_{E-inertia}$: $\lambda s.\lambda s'.$ .s′ is an Event-inertia situation for s, where for any two situations s and s′, s′ is an Event-inertia situation for s iff all the events that have actually started in s continue in s′ as they would if there were no interruptions.

4 Incorporating tense and aspect information in semantic representations

In this section I provide a concrete implementation of the semantic properties outlined above. This includes both representations for verbal aspect and viewpoint aspect. The implementation is an extension of an existing semantic formalism for LFG, namely abstract knowledge representations (Bobrow et al. 2007). AKR is based on textual inference logic (Bobrow et al. 2005) and employs two different layers of meaning. The first is a conceptual layer that is inspired by Neo-Davidsonian event semantics De Paiva et al. (2007). Thus, the conceptual structure contains information about the predicate-argument structure on the one hand, but also serves as the interface to an underlying ontology in this case based on WordNet and VerbNet data. The `subconcept` facts in the example below introduce the lexical elements and their references to the ontology. The ontology consists of enumerated word senses with corresponding synsets containing information, e.g., on hypernyms, which is required for corresponding reasoning tasks. The subconcept introducing the main predicate, `draw:9` can be understood roughly as an the event variable, while the other subconcepts introduce elements that correspond to other types of entities. However, it is important to note that the elements of the conceptual structure have no extensions in the actual world. Thus, the name *conceptual structure*.

```
Conceptual Structure:
    definite(John:2)
    subconcept(draw:9,[pull-1,reap-2,trace-2,draw-4,...])
    role(Theme,draw:9,circle:19)
    role(Agent,draw:9,John:2)
    subconcept(John:2,[male-2])
    role(cardinality_restriction,John:2,sg)
    subconcept(circle:19,[circle-1, set-5,...])
    role(cardinality_restriction,circle:19,sg)
```

Figure 2: Conceptual structure: *John is drawing a circle.*

The subconcepts of the conceptual structure are instantiated in so-called contexts. Facts about this are stored in the contextual structure. This roughly corresponds to acknowledging an extension of the corresponding object in the actual world (roughly the top context) or some possible world.

```
Contextual Structure:
        context(t)
        top_context(t)
        instantiable(John:2,t)
        instantiable(circle:19,t)
        instantiable(draw:9,t)
```

Figure 3: Contextual structure: *John is drawing a circle.*

Embedded contexts are introduced by modal elements, e.g., modal verbs or propositional attitude verbs. Consider what happens when we embed the sentence above in the complement of the propositional attitude verb *know*. In this case the complementizer *know* introduces a new context anchored to its theme argument *draw*. This new context commits to a possible situation in which John draws a circle by means of the instantiation facts. However, the attentive reader will realize that the semantic representation commits to instantiation of the situation in the top context. This is a result of the factivity presupposition (Karttunen 1971) of the propositional attitude verb *to know*, indicated by the *veridical* context lifting relation in the contextual structure. In other words, by virtue of the factivity presupposition, a *veridical* context lifting relation raises the content of the embedded context to the content of the matrix context (Bobrow et al. 2007).

```
Conceptual structure:
        definite(Maria:1)
        subconcept(know:7,[know-1,...])
        role(Theme,know:7,ctx(draw:9))
        role(Agent,know:7,Maria:1)
        subconcept(Maria:1,[female-2])
        role(cardinality_restriction,Maria:1,sg)
Contextual Structure:
        context(t)
        context(ctx(draw:9))
        top_context(t)
        context_lifting_relation(veridical,t,ctx(draw:9))
        context_relation(t,ctx(draw:9),crel(Theme,know:7))
        instantiable(John:2,t)
        instantiable(Maria:1,t)
        instantiable(circle:19,t)
        instantiable(draw:9,t)
        instantiable(know:7,t)
        instantiable(John:18,ctx(draw:9))
        instantiable(circle:19,ctx(draw:9))
        instantiable(draw:9,ctx(draw:9))
```

Figure 4: AKR: *Maria believes that John is drawing a circle.*

An *averdical* context relation does not introduce a context lifting relation. For example embedded under the verb *believe*, the facts that constitute a situation where John is drawing a circle would remain only in the embedded context. This affects the reasoning about factive vs. non-factive verb. More on that in Section 4.3.

An *antiveridical* context relation is for example introduced by negation. Antiveridicality introduces uninstantiability. It is noteworthy, that only the "event" is `uninstantiable` in Figure 5. This is desired, since negation of an event only subtracts those situations from the top context where the event did not take place, whereas it does not subtract those situations where its thematic arguments exist. This is important since I use a similar line of reasoning to account for perfective vs. imperfective interpretations below.

```
Contextual Structure:
        context(t)
        context(ctx(draw:13))
        top_context(t)
        context_lifting_relation(antiveridical ,t,ctx(draw:13))
        context_relation(t,ctx(draw:13),not:10)
        instantiable(John:1,t)
        uninstantiable(draw:13,t)
        instantiable(circle:20,ctx(draw:13))
        instantiable(draw:13,ctx(draw:13))
```

Figure 5: Contextual structure: *John did not draw a circle.*

For reasons of space, I simplify the structures above to simpler graph structures.[5] I remove some unnecessary information for the present purposes: I omit semantic and ontological properties from the conceptual structure where they can be derived from the example sentence. The label `top_context` is omitted from the contextual structure. Thus, the representation for *John is drawing a circle* is the following:

Figure 6: AKR: *John drew a circle.*

Veridical context lifting relations between contexts are introduced by continuous lines, while antiveridical relations are introduced by dashed lines. This is illustrated in Figure 7 by means of the sentence *John did not draw a circle*. Uninstantiable nodes are marked with `not(...)`.

[5]This idea is inspired by the actual graph-based knowledge representations by Kalouli & Crouch (2018). However, the graphs presented here are strict translations of AKR not GKR as presented in the cited work.

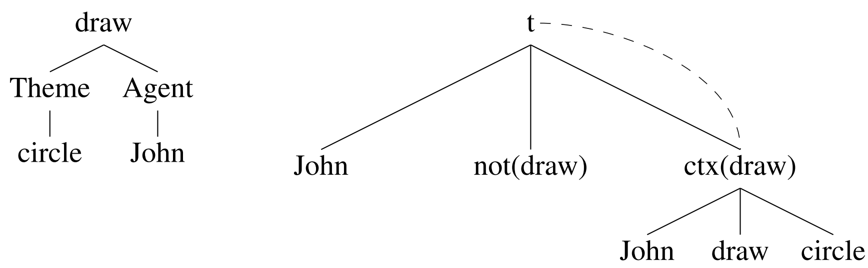


Figure 7: AKR: *John did not draw a circle.*

4.1 Enriching the Contextual Structure

The glaring issue with AKR that this paper provides a solution for is the fact that the AKR in Figure 6 does not express any temporal or aspectual information at all. This paper focusses on the analysis of aspectual properties. Temporal relations are assumed to be stored in a separate structure and to hold between contexts as well as temporal intervals. I subsume all of these elements under the label of situation in the sense of a world/time pair.[6] Temporal information will be displayed where necessary.

The goal of this paper is to change the contextual structure in AKR. Until now, the approach has been fleshed out in such a way that contexts are closer to possible worlds rather than situations in the sense that no temporal information is incorporated in them.[7] Thus, the AKR in Figure 6 can be mapped on any of these following sentences:[8]

(17) a. John draws a circle.
b. John drew a circle.
c. John is drawing a circle.
d. John was drawing a circle.
e. John has drawn a circle.

I propose a more fine grained representation that (for the sake of this paper) assumes the perfective representation as the default (see Figure 8) and the progressive is represented in a more detailed structure (see Figure 9). Both the progressive and habitual in Figure 10 introduce an intensional context. They, thus, correspond roughly to the intensional quantifiers discussed in Section 3. More concretely, the *progressive* and the *habitual* describe different relations between the `top_context` and the embedded `ctx`. The latter can be treated in the same vain as any other averidical relation, e.g., those provided by non-factive attitude verbs. The treatment of the *progressive* on the other hand is novel, yet simple. The idea is, that the whole eventuality and its arguments is embedded intensionally in an averdical relation to the `top_context`. However, the conceptual description

[6]Theoretically this approach is closer to Barwise & Perry (1981).
[7]At least without concrete temporal modifiers such as *yesterday* or *tomorrow*.
[8]Based by the entailment pattern provided by the system described in Bobrow et al. (2007).

of the event itself is raised to the `top_context`. It thus shares the conceptual structure with its embedded counterpart but behaves differently in terms of intensional reasoning. This is explained in detail in Section 4.3. In the next section, I illustrate the difference between *events* and *processes* at the level of the conceptual structure and how it affects the corresponding contextual structure discussed in this section.

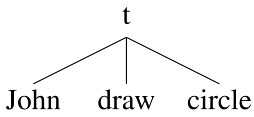


Figure 8: AKR: *John drew a circle.*

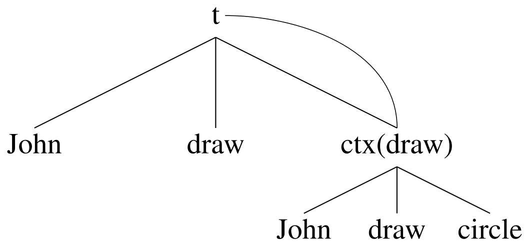


Figure 9: AKR: *John is/was drawing a circle.*

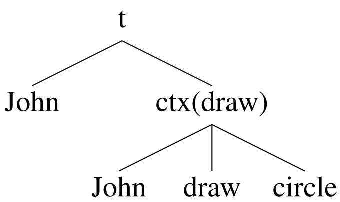


Figure 10: AKR: *John draws circles.*

4.2 Enriching the Conceptual Structure

The conceptual structure of an AKR is created from various lexical resources, in particular, WordNet and VerbNet. Those resources are (partially) compiled into a unified lexicon (Crouch & King 2005). The unified lexicon disambiguates word senses on the basis of f-structure correspondences where possible.[9]

In the previous section, I explained that the imperfective paradox arises due to the different event classes that it can apply to, i.e., events and processes. Unfortunately, this information is not straightforwardly derivable from the conceptual structure in the AKR. To remedy this, I propose to provide an extension of the conceptual structure based on decompositional event semantics. I illustrate this in terms of an an implementation of Ramchand's (2008) first-phase syntax (FPS). The template for the FPS is shown in Figure 11.

[9]The unified lexicion ascribes word senses to certain grammatical structures, e.g., `V-SUBJ-OBJ`). These are aligned with the subcategorization frame employed by the syntactic component. Even then the output is often highly ambiguous since multiple word senses might be associated with the same subcat frame. For reasons of space I ignore this issue for the most part.

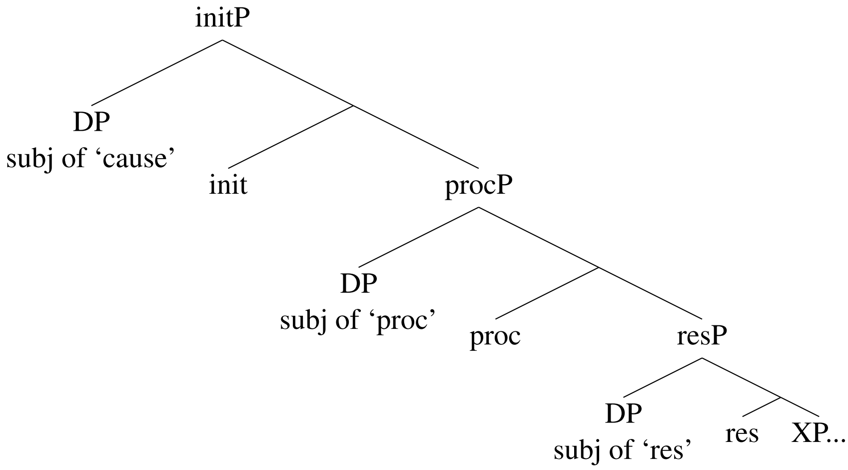


Figure 11: First-phase-syntax template (Ramchand 2008)

As the template illustrates, the FPS assumes three different syntactic heads containing two different elements. First consider the stages of a given eventuality: *init*, *proc*, and *res*. *init* and *res* are states that refer to the state before and after the change described by *proc*.[10] Second, the *subj* slots are provided for the corresponding thematic roles. The *subj of 'cause'* can be derived mnemonically from the corresponding head: it denotes the iniator. The *subj of 'proc'* is generally an *undergoer* and the *subj of 'res'* is the resultee. An overview of correspondences between the different sub events and the corresponding thematic roles can be found in Ramchand (2008: 98). The correspondences between FPS and aspectual class as presented above are given in (18). An eventuality, either a process or an event, is a sequence of sub-eventualities written as $< \phi, \psi, ... >$. Parenthesis denote an optional object.[11] For the present purposes a *telic argument* is quantized or describes a (finite) path. Conversely, an atelic argument is cumulative, or does not otherwise specify an endpoint (Krifka 1998).

(18) a. *Event* $\rightarrow < (init), proc, res >$
Process + *"telic argument"*
b. *Process* $\rightarrow < (init), proc >$ + *"(atelic argument)"*

This means that the syntactic structure inside *V* is as in Figure 12 for the two examples relevant to the imperfective paradox.

[10]The boundaries are not completely rigid, in fact there can be a complete overlap between *init* and *process* (Ramchand 2008). Thus, I take it that *init* includes a preparatory stage in the sense of Moens & Steedman (1988). And *proc* refers to the actual event taking place.

[11]Optional initiators are for example found in sentences like *the snow melted*, where *melt* describes a process without an active initiator (*subj of 'cause'*).

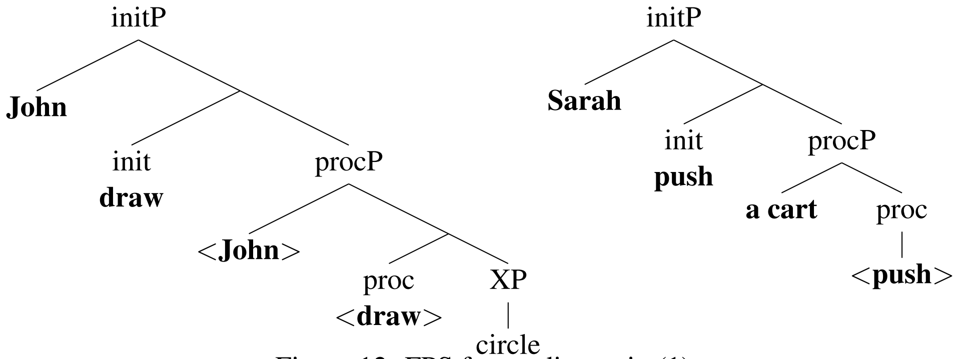


Figure 12: FPS for predicates in (1)

Given this additional structure we can define a procedure that maps decomposed events onto contextual structures as presented in Section 4.1. However, an additional step is required since neither of the predicates *draw* and *push* has a syntactically realized result state in the sense of Ramchand & Svenonius (2014). The *XP* in Figure 12 denotes an *telic argument*. Similar, to proper result states we can derive a situation that expresses the result state of *drawing a circle*, namely, the situation in which a circle has come into existence. This can be modelled in terms of *instantiability* in the case of both consumption and creation verbs. An object coming into existence becomes instantiable, and an object that is consumed becomes uninstantiable.

In the case of the creation verb *draw*, the result state is modeled as a context `ctx` in which *circle* is instantiated. By simply mapping all unique elements on to a context we achieve our default perfective interpretation which entails the achievement of the result state. From this we raise the fact that corresponds to the *procP* in the FPS to the `top_context`, following the intuition that the progressive is derived from its resultative (perfective) form we have commited to in Section 3.

The derivation of the result state has to be adapted for different possible *XP* objects such as consumables or paths. In particular paths as, e.g., in *Sarah pushed the cart to the store* provide a challenge. This is due to the fact that there is neither a conceptual nor a contextual description that corresponds to what it means to follow a path.[12] However, leaving the issue underspecified might work well enough. The perfective representation could simply entail proximity to the goal of the path, where as in the progressive variant the `top_context` which does not instantiate the *goal* corresponds to the situation in which there is no proximity between *initiator*, *undergoer* and *goal*. The specifics for various verb classes are left for future research.

[12] Zaenen et al. (2008) provide an approach to predicates of change-of-location in AKR and will be considered in future work. Thanks to a reviewer for pointing this out to us.

4.3 Inferencing

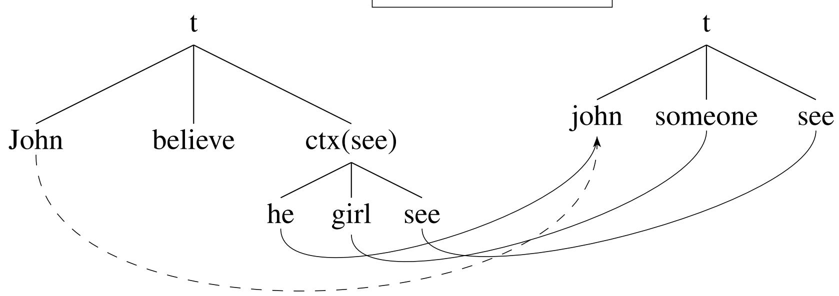
AKRs are specifically designed for inference tasks. In the next few paragraphs, I give a rough sketch of the inferencing techniques employed in combination with AKR representations (see Bobrow et al. (2007) for a more extensive explanation).

AKR uses a so-called entailment and contradiction detection (ECD) system for detecting inferences. As the name suggests it tests whether an entailment or contradiction relation holds between two linguistic expressions. The basic idea is fairly straightforward: First, the premise and the conclusion are aligned on a conceptual level. If two concepts conflict (e.g., due to negation in *a boy smiled* → *a boy did not smile*), a contradiction flag is introduced. A successful alignment is shown in (19). The concepts *John* and *saw* are identical between the sentences and thus alignment is trivial. The specific concept *girl* entails the less specific concept *someone* but not vice versa. Thus, the inference in (19) is valid, but the reverse inference is not.

(19) John saw a girl → John saw someone `YES`

The ECD not only checks for conceptual similarity but also checks for intensional reasoning by means of the contextual structure. Consider the example below. Figure (20) illustrates that, conceptually, an alignment similar to the one in (19) can be found. However, this time the premise has instantiated the concepts in an intensional context (introduced through the non-factive verb *believe*). Since non-factive verbs do not introduce a context-lifting relation, the concepts are instantiated only in the hypothetical context `ctx(see)`. Although the concept *John* is instantiated in the top context the alignment is not fruitful as represented by a dashed line.

(20) John believes that he saw a girl → John saw someone `UNKNOWN`

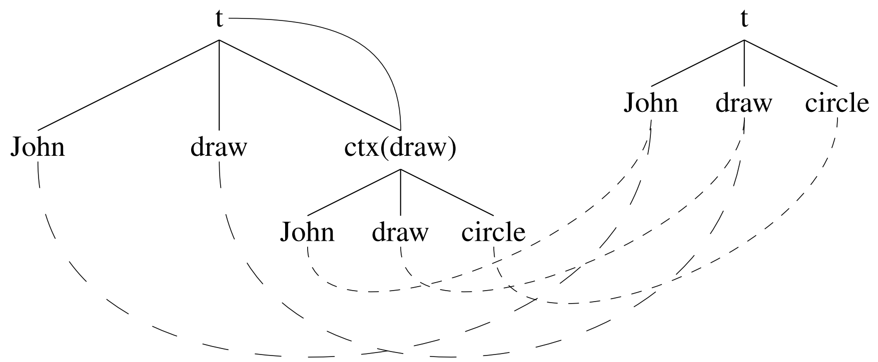


Formally, the system calculates inferences by deleting aligned concepts if their contexts are compatible. This means, in (19) all concepts are deleted since they belong to the same context. The result is an entailment relation. In (20) the system will not delete the aligned facts since they make commitments about different contexts. The inference in (20) is not valid. Contradictions are introduced by specific flags. As this explanation suggests, ECD has three possible output states: `YES` for entailments, `NO` for contradictions, and `UNKNOWN` for all other cases.

4.4 Reasoning with aspect

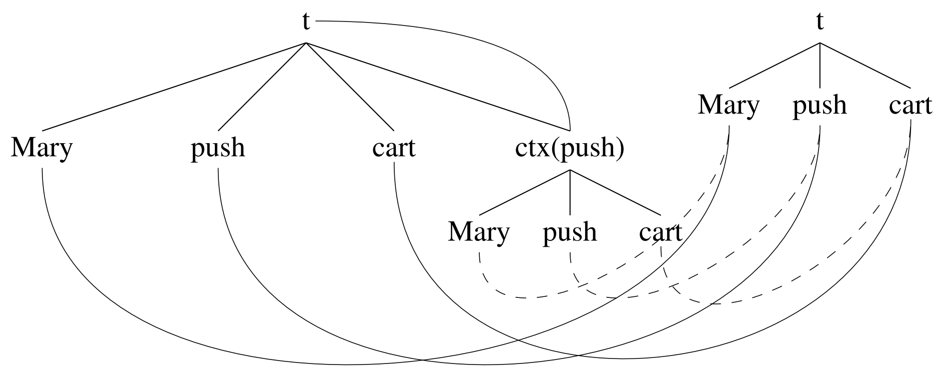
Example (21) illustrates why the inference from an event in the progressive form such as *(to) be drawing a circle* to its simple past counter part fails: The conclusion in (21) may be fully aligned with the intensional context introduced by the progressive. However, due to the mismatch between `top_context` and embedded context, this alignment does not lead to the deletion of the corresponding conclusion facts. Furthermore, the conclusion can only be partially aligned with the top context of the premise. Thus, based on this alignment the corresponding facts in the conclusion will not be fully deleted either. The fact corresponding to the *circle* remain.

(21) John was drawing a circle → John drew a circle `UNKNOWN`



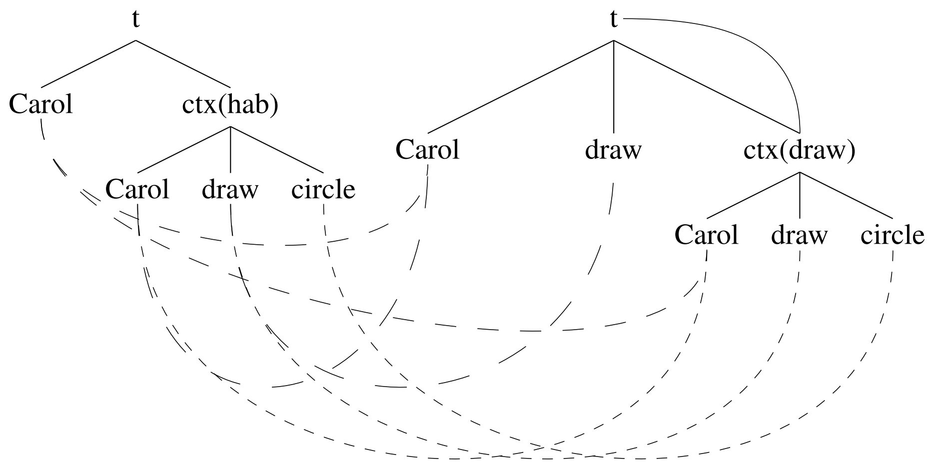
Compare this to example (22). Applied to an achievement the progressive form provides a different `top_context`. Now, the conclusion is fully entailed in the premise conceptually as well as contextually, which leads to entailment detection.

(22) Mary was pushing a cart → Mary pushed a cart `YES`



The third kind of inference, which I mentioned in the beginning of the paper, is the inference from a habitual to its progressive counterpart, as shown in (23). The inference should fail as indicated. The corresponding contextual alignment seems messy but the inference ultimately fails simply due to the instantiation of *draw* in the `top_context` in the conclusion, which does not find a corresponding element in the `top_context` of the antecedent.

(23) Carol draws circles → Carol is drawing a circle `UNKOWN`



A challenge for the approach presented here is presented in (24). I established before, that the progressive does not commit to the fact that the corresponding event reaches its intended endpoint. However, the circumstances can be modified, for example, by the conjunct added in the premise in (24) to the effect that the intended endpoint and consequently the result state can never be reached.

(24) *The chicken was crossing the road*, when it got hit by a truck → *The chicken crossed the road* `UNKNOWN`

However, I argue that this is not a problem of the present approach but rather an issue of incorporating world knowledge into the inference system. In fact, whether the inference is valid or not is still unknown because it is not clear whether the accident simply changes the path of the chicken or actually prevents it from achieving its goal.

5 Conclusion

In this paper I have laid the groundwork for aspectual reasoning with help of computational LFG methods. I draw from existing resources, in particular the bridge system developed at PARC (Bobrow et al. 2007). By introducing additional layers of semantic information, such as temporal and aspectual information, the contextual structure of AKRs has been made substantially more expressive. The change can be understood as a movement from contexts as possible worlds to contexts as world/time pairs or situations.

This system works under the precondition that the semantic representation (based on Neo-Davidsonian event semantics), from which an AKR is derived, is enriched in terms of decompositional event semantics à la Ramchand (2008). This information together with the information provided by the lexical resources incorporated in the PARC bridge system allows for the implementation of a principled way for treating tense and aspect that captures the basic inference patterns observed in the literature.

References

Arregui, Ana, María Luisa Rivero & Andrés Salanova. 2014. Cross-linguistic variation in imperfectivity. *Natural Language & Linguistic Theory* 32(2). 307–362.

Asher, Nicholas. 1992. A Default, Truth Conditional Semantics for the Progressive. *Linguistics and Philosophy* 15(5). 463–508.

Barwise, Jon & John Perry. 1981. Semantic Innocence and Uncompromising Situations. *Midwest Studies in Philosophy* 6(1). 387–404.

Bertinetto, Pier Marco & Alessandro Lenci. 2012. Habituality, Pluractionality and Imperfectivity .

Bobrow, Daniel G., Bob Cheslow, Cleo Condoravdi, Lauri Karttunen, Tracy Holloway King, Rowan Nairn, Valeria de Paiva, Charlotte Price & Annie Zaenen. 2007. PARC's Bridge and Question Answering System. In *Proceedings of the GEAF 2007 Workshop*, 1–22.

Bobrow, Danny, Cleo Condoravdi, Richard Crouch, Ronald Kaplan, Lauri Karttunen, Tracy Holloway King, Valeria de Paiva & Annie Zaenen. 2005. A Basic Logic for Textual Inference. In *Proceedings of the aaai workshop on inference for textual question answering*, 47–51.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The Parallel Grammar Project. In *Proceedings of the 2002 Workshop on Grammar Engineering and Evaluation*, vol. 15, 1–7. Association for Computational Linguistics.

Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. CSLI Publications.

Butt, Miriam, María-Eugenia Niño & Frédérique Segond. 1996. Multilingual Processing of Auxiliaries within LFG. In *Konvens*, vol. 96, 111–122.

Carlson, Greg. 2012. Habitual and Generic Aspect. In Robert I. Binnick (ed.), *The Oxford handbook of Tense and Aspect*, Oxford University Press.

Cipria, Alicia & Craige Roberts. 2000. Spanish imperfecto and Pretérito: Truth Conditions and Aktionsart Effects in a Situation Semantics. *Natural Language Semantics* 8(4). 297–347.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*, vol. 2. Cambridge university press.

Crouch, Dick, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2017. *XLE Documentation*. Palo Alto Research Center.

Crouch, Dick & Tracy Holloway King. 2005. Unifying Lexical Resources. In *Proceedings of the interdisciplinary workshop on the identification and representation of verb features and verb classes*, 32–37.

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34 Syntax and Semantics. New York: Academic Press.

De Paiva, Valeria, Daniel G. Bobrow, Cleo Condoravdi, Richard Crouch, Lauri Karttunen, Tracy Holloway King, Rowan Nairn & Annie Zaenen. 2007. Textual Inference Logic: Take Two. *Contexts and Ontologies Representation and Reasoning* 32.

De Swart, Henriette. 1998. Aspect Shift and Coercion. *Natural Language & Linguistic Theory* 16(2). 347–385.

Depraetere, Ilse. 1995. On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy* 18(1). 1–19. doi: 10.1007/BF00984959. http://dx.doi.org/10.1007/BF00984959.

Dowty, David R. 1977. Toward a Semantic Analysis of Verb Aspect and the English 'Imperfective' Progressive. *Linguistics and philosophy* 1(1). 45–77.

Fara, Michael. 2005. Dispositions and habituals. *Noûs* 39(1). 43–82. http://www.jstor.org/stable/3506173.

Ferreira, Marcelo. 2016. The Semantic Ingredients of Imperfectivity in Progressives, Habituals, and Counterfactuals. *Natural Language Semantics* 24(4). 353–397.

Filip, Hana. 2012. Lexical aspect. *The Oxford handbook of tense and aspect* 721–751.

Ide, Nancy & James Pustejovsky. 2017. *Handbook of Linguistic Annotation*. Springer Netherlands.

Kalouli, Aikaterini-Lida & Richard Crouch. 2018. GKR: the Graphical Knowledge Representation for semantic parsing. In *Proceedings of the Workshop on Computational Semantics beyond Events and Roles*, 27–37.

Karttunen, Lauri. 1971. Some Observations on Factivity. *Research on Language & Social Interaction* 4(1). 55–69.

Klein, Wolfgang. 1994. *Time in Language*. Psychology Press.

Kratzer, Angelika et al. 1991. Modality. *Semantics: An International Handbook of Contemporary Research* 7. 639–650.

Krifka, Manfred. 1998. The Origins of Telicity. In *Events and Grammar*, 197–235. Springer.

Landman, Fred. 1992. The Progressive. *Natural Language Semantics* 1(1). 1–32.

Löbner, Sebastian. 2002. Is the German Perfekt a Perfect Perfect. In Ingrid Kaufmann & Barbara Stiebels (eds.), *More than Words: A Festschrift for Dieter Wunderlich*, Akademie Verlag.

Moens, Marc & Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14(2). 15–28.

Musan, Renate. 2002. *The German Perfect: Its Semantic Composition and its Interactions with Temporal Adverbials*, vol. 78. Springer Science & Business Media.

Parsons, Terence. 1990. *Events in the Semantics of English*, vol. 5. MIT Press.

Portner, Paul. 2019. Perfect and Progressive. *Semantics-Noun Phrases and Verb Phrases* 313.

Ramchand, Gillian & Peter Svenonius. 2014. Deriving the Functional Hierarchy. *Language Sciences* 46. 152–174.

Ramchand, Gillian Catriona. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*, vol. 116. Cambridge University Press.

Reichenbach, Hans. 1947. The Tenses of Verbs. *Elements of Symbolic Logic* 287–298.

Smith, Carlotta. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

Steedman, Mark. 1997. Chapter 16 - Temporality. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of Logic and Language*, 895 – 938. North-Holland.

de Swart, Henriëtte. 2016. Aspectual Composition and Recursion. *Mood, Aspect, Modality Revisited: New Answers to Old Questions* 314.

Vendler, Zeno. 1957. Verbs and times. *The philosophical review* 66(2). 143–160.

Zaenen, Annie, Daniel Bobrow & Cleo Condoravdi. 2008. The Encoding of Lexical Implications in VerbNet Predicates of Change of Locations. In Nicoletta Calzolari et al. (ed.), *LREC'08 proceedings*, ELRA.

Zymla, Mark-Matthias. 2017a. Comprehensive Annotation of Cross-Linguistic Variation in the Category of Tense. In *12th International Conference on Computational Semantics*, http://www.aclweb.org/anthology/W17-6817.

Zymla, Mark-Matthias. 2017b. Cross-Linguistically Viable Treatment of Tense and Aspect in Parallel Grammar Development. In *Proceedings of the LFG17 Conference*, CSLI Publications.

On the Syntax/Semantics Interface in Computational Glue Semantics: A Case Study

Mark-Matthias Zymla
University of Konstanz

Gloria Sigwarth
University of Konstanz

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 374–392

http://csli-publications.stanford.edu/LFG/2019

Keywords: glue semantics, workbench, sorted type theory

Zymla, Mark-Matthias, & Sigwarth, Gloria. 2019. On the Syntax/Semantics Interface in Computational Glue Semantics: A Case Study. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 374–392. Stanford, CA: CSLI Publications.

Abstract

This paper describes an extension of the Glue semantics workbench by Meßmer & Zymla (2018). In particular, we present a version of the workbench that can deal with (at least) semantic formalisms based on (Two-)Sorted Type Theory. We illustrate this by providing a semantic analysis of different constructions that involve quantification over the verb: adverbs, raising verbs, control verbs and attitude verbs. Furthermore, we describe some additional features of the workbench that aim at improving the workflow with the system.

1 Introduction

This paper expands on the effort initiated by Meßmer & Zymla (2018) to provide a (new) computational implementation of Glue semantics (Dalrymple 2001). This implementation is called Glue semantics workbench (GSWB) and provides a linear logic prover based on an algorithm by Hepple (1996) and an implementation of lambda calculus within a Java program.

We present a user study for the Glue semantics workbench and report on various issues that have been detected and fixed. This study is based on a situation semantics implementation as presented in Kallmeyer & Romero (2008) for LTAG. Their paper discusses several constructions that have been explored in the LFG literature from a resource logic perspective: adverbs, raising verbs, control verbs and attitude verbs etc. Example (1) illustrates some expressions we use as working examples for this paper. We follow the line of work presented in Asudeh (2005, 2000, 2002) and reconcile it with the aforementioned situation semantics approach. The goal thereby is to achieve an appropriate treatment of the scope interactions shown to the right in (1).

(1)	a.	John sometimes laughs.	
	b.	Every girl sometimes laughs.	$\forall > \exists, \exists > \forall$
	c.	John sometimes kisses every girl.	$\forall > \exists, \exists > \forall$
	d.	Paul claims Mary apparently loves John.	$claims > apparently$

The case study presented here illustrates how the GSWB works for different typed semantic systems. The insights described in this paper thus provide a broader perspective on the use of the GSWB in general.

The paper is structured as follows: in Section 2 we discuss the theoretical background of situation semantics and the constructions presented in (1). In Section 3.2 we illustrate how we provide derivations for these expressions with the help of the GSWB. In Section 3.3 we present further smaller additions to the workbench the need for which arose during this case study. Section 4 concludes this paper.

† We thank the *VALIDA* project for funding. Furthermore, we thank Maribel Romero and Miriam Butt for helpful discussion. Finally, we thank the anonymous reviewers as well as the audience of the LFG2019 conference.

2 Theoretical background

The original GSWB by Meßmer & Zymla (2018) provided a Glue semantics prover and a simple syntax/semantics interface with the help of which basic compositional issues such as the quantifier interactions in (2) and adjectival modification could be illustrated.

(2) Every man loves a woman.

a. $\forall x[man(x) \rightarrow \exists y[woman(y) \wedge love(x, y)]]$

b. $\exists x[woman(x) \wedge \forall y[man(y) \rightarrow love(x, y)]]$

In this paper, we presuppose a basic understanding of Glue semantics that underlies these kinds of constructions. For reference we refer to Meßmer & Zymla (2018), Dalrymple et al. (1999) and Crouch & van Genabith (2000). Our main goal is to move from quantification and modification in the nominal domain to quantification in the domain of verbs. The primary observation, in this domain, is that quantification scope is much more restricted. More specifically, the semantics of verbal quantifiers are more strictly intertwined with the verb element they scope over (Cinque 1999). This is dubbed by Kallmeyer & Romero (2008) the issue of *Quantification at the verbal spine*. Their leading point of discussion is provided by the examples below in (3) and (4). The former illustrates the "rigid scope of (ad)verbal attachments" (Kallmeyer & Romero 2008). (4) on the other hand illustrates the flexible scope of NP quantifiers in relation to verb quantification.

(3) John seems to sometimes laugh.

a. seem(sometimes(laugh(j)))

b. *sometimes(seem(laugh(j)))

(4) John seems to have visited everybody $\quad seem > \forall, \forall > seem$

Kallmeyer & Romero (2008) proceed to discuss the scope restrictions for both (ad)verbal elements as well as NP quantifiers. Examples (5) and (6) (still by Kallmeyer & Romero (2008)) highlight the role of finite clause boundaries. They state the assumption that NP quantifiers are limited to the first finite clause containing the NP.

(5) A student wants to meet every professor $\quad \exists > \forall, \forall > \exists$

(6) A student said that you met every professor $\quad \exists > \forall, *\forall > \exists$

This paper tests some hypotheses made by Kallmeyer & Romero (2008) as well as researchers within the glue community (e.g. Asudeh (2005), Dalrymple et al. (1999)). However, we deviate slightly from the typical practice of assuming an underlying event semantic framework. This is explained in the next section.

2.1 Situation semantics

In this paper we opt for an approach that uses situation variables of type s in the object language (i.e. the meaning side in linear logic), next to individuals of type e. According to Kratzer (2019), situations in natural language semantics date back to Barwise (1981), where they were used for the semantics of perception reports. These days, situation semantics is an alternative, or sometimes an extension, to possible world semantics. Intuitively, linguistic expressions are evaluated with regard to partial worlds and not to complete worlds.

The use of situation semantics departs from the approaches presented in much previous work on the kind of constructions discussed in this paper. See, e.g., Asudeh (2000, 2002, 2005) for control and raising. On the other hand we take inspiration from works such as Haug (2008) and Lowe (2014) who deal with phenomena such as tense and aspect. Following Kallmeyer & Romero (2008), the chosen meaning language follows Two-Sorted Type Theory, meaning that each predicate has a situation argument in its semantic denotation (see (7-a)).

We believe that working with these more complex structures allows us to generalize to simpler structures.[1]. In fact, many examples given in the works referred to here have been tested in the GSWB, although Haug (2008) has been slightly modified since the GSWB is not able to deal with compound types yet. This is due to the lack of multiplicative conjunction in the employed linear logic fragment.

In the present paper, the situation variable is assumed to be anchored to the TNS-ASP node and is, thus, taken to be part of the verb meaning constructor. This is illustrated in (7-a).[2]

(7) a. $[\![\, \textit{laughs} \,]\!] = \lambda x_e.\lambda s_s.laughs(x, s) : g \multimap (h \multimap f)$
$[\![\, \textit{Tina} \,]\!] = t : g$
$[\![\, \exists\textit{-closure} \,]\!] = \lambda p_{st}.\exists s[p(s)] : (h \multimap f) \multimap f$
b. $[\![\, \textit{Tina laughs} \,]\!] = \exists s[laughs(t, s)] : f$

Especially, in the domain of control and raising verbs this opens up questions about the role of inflection on the embedded verb. This issue will be addressed briefly when reviewing Asudeh's proposals for these kinds of constructions in the light of the present paper in section 2.4.

2.2 Scope restrictions in Glue semantics

The advantage of using Glue semantics as a semantic formalism is, that many scope interactions fall out naturally as a consequence of the derivation system. This has long since been shown by Dalrymple et al. (1999) and was the first milestone of the GSWB. However, Gotham (2019) discusses flexible and rigid scope configurations of NPs and their treatment in LFG + Glue (i.e. Glue Semantics for Lexical

[1]This is not meant as a critique of the more simple structures employed in some papers since they aim at resolving issues for which more complex semantics would needlessly complicate the analysis.

[2]In the derivations to come, we will usually omit the step of existential closure.

Functional Grammar). Furthermore, many more actual and apparent scope issues still remain unsolved, e.g. the role of definites (see Heim (2011) for an overview) as well as a comprehensive overview over the quantification of verbs, which takes center stage in this paper.

Similarly to Asudeh (2005) (in the domain of control and raising verbs), we aim at semantics of verb quantification, where the scope interactions are restricted naturally by the linear logic side of Glue semantics meaning constructors. This is mainly due to the fact, that we want to keep to the linear logic fragment employed in the GSWB as is.

To begin with, let us clarify the general semantics we envision for the types of verbs discussed in this paper. We base our work on Hintikka-style semantics (Hintikka 1969). An example of this is given in (8).[3]

(8) $[\![\, \textit{seem} \,]\!] = \lambda p_{<s,t>}.\lambda s_0.\forall s'[s' \in SEEM(s_0) \rightarrow p(s')]$

Example (8) is a standard case of our treatment of adverbs, control and raising verbs, as well as attitude verbs, as universal quantifiers over situations in the semantics.[4] The different kinds of quantifiers are distinguished by different restrictors over the situations they bind. Although we propose a fairly similar semantics for all the phenomena we discuss in this paper, as is to be expected, there will emerge subtle differences in the following sections of this paper. Some of these are influenced by the semantic side, e.g. the choice whether propositions or properties are embedded under a given kind of verb; others help to guarantee certain scope configurations. This will be done mainly by appropriate anchoring of situation variables to the f-structure and the use of linear logic variables. The next section on adverbs will illustrate this procedure.

2.3 Adverbs

In this section we discuss adverbs, in particular, those that attach directly to the verb. In computational LFG (i.e. in the English XLE grammar) adverbs are ambiguous between a sentence scope (`sadv`) reading and a verb scope reading (`vadv`). This ambiguity overlaps with the distinction between frequency adverbs and adverbs of quantification. In this paper, we focus on the latter. From an XLE perspective, we attribute quantificational adverbs or quantificational readings of adverbs to the category `vadv`.[5]

We treat adverbs as simple modifiers. Modifier premises have a particular structure on their glue side: A linear implication with an equivalent antecedent and consequent. In the case we discuss below, they take propositions as their ar-

[3]The notation $SEEM(s_0)$ stands for the set $\{s'$: s' conforms to what appears to be the case in $s_0\}$

[4]Not all quantifiers receive a Hintikka-style modal semantics, but they still follow the general template of universal quantification.

[5]The treatment of sentence adverbs and frequency adverbials (Bennett & Partee 1978) and the exact distinction between frequency adverbs and quantificational adverbs is left for future work.

gument. This is illustrated in (9). Although the adverb is attached directly to the top-level verb by means of the reference to ↑`TNS-ASP` as well as ↑, the sentence has two readings. This is, of course, due to fact that the quantificational NP can scope under (surface scope) or over the adverb (inverse scope). (10) presents the appropriate lexical entries. The Glue formulas are demonstrated first in their general form and secondly, the resources are instantiated to nodes in the f-structure in Figure 1.[6]

$$
f\left[\begin{array}{lll}
\text{PRED 'kiss<John,girl>'} & & \\
\text{ADJUNCT} & \left\{\begin{array}{l}\text{PRED 'sometimes'}\\ \text{ADV-TYPE vpadv}\end{array}\right\} & \\
\text{SUBJ} & g\left[\begin{array}{ll}\text{PRED 'John'} & \\ \text{NTYPE} & \left[\text{NSYN proper}\right] \\ \text{PERS 3, NUM sg} & \end{array}\right] & \\
\text{OBJ} & k\left[\begin{array}{ll}\text{PRED 'girl'} & \\ \text{SPEC} & \left[\text{PRED 'every'}\right] \\ \text{NTYPE} & \left[\text{NSYN common}\right] \\ \text{PERS 3, NUM sg} & \end{array}\right] & \\
\text{TNS-ASP} & t\left[\begin{array}{l}\text{TENSE pres, PROG -\_, PERF -\_,}\\ \text{MOOD indicative}\end{array}\right] &
\end{array}\right]
$$

Figure 1: *John sometimes kisses every girl.*

(9) John sometimes kisses every girl.

(10) a. $[\![$ *John* $]\!] = j : g$

b. $[\![$ *sometimes* $]\!] = \lambda p_{<s,t>}.\lambda s_0.\exists s[s \leq s_0 \wedge p(s)]$:
$(\uparrow \text{TNS-ASP} \multimap \uparrow) \multimap (\uparrow \text{TNS-ASP} \multimap \uparrow)$
$= (t \multimap f) \multimap (t \multimap f)$

c. $[\![$ *kiss* $]\!] = \lambda y_e.\lambda x_e.\lambda s.kiss(x,y,s)$:
$\uparrow \text{OBJ} \multimap (\uparrow \text{SUBJ} \multimap (\uparrow \text{TNS-ASP} \multimap \uparrow))$
$= k \multimap (g \multimap (t \multimap f))$

d. $[\![$ *every girl* $]\!] = \lambda Q_{<e,<s,t>>}.\lambda s.\forall x[girl(x,s) \rightarrow Q(x)(s)]$:
$\forall X, Y.((\uparrow \text{OBJ} \multimap (Y \multimap X)) \multimap (Y \multimap X))$
$= \forall X, Y.((k \multimap (Y \multimap X)) \multimap (Y \multimap X))$

[6]Common nouns also have situation variables as can be inferred from the denotation of the quantifier *every girl*. It is less clear how to anchor this situation variable in the f-structure. If need be for anchoring, it would probably be somewhere in the `NTYPE` grammatical feature that specifies various semantic features in f-structure. `NTYPE` is used uniformly across the ParGram XLE grammars (Butt et al. 2002).

Note that the quantifier now maps onto elements of type $< s, t >$ both of which are treated as variables in its denotation. This is how we model the more flexible nature of NP quantifiers. In contrast to the quantificational NP, there are no Glue variables needed in the meaning constructor for *sometimes*, because the adverb takes scope where it attaches.

Using the meaning constructors in (10), the Glue proofs for both readings of (9) can be constructed. The proofs are shown in Figures 2 and 3. For readability, the semantic side is omitted.

$$\dfrac{(t \multimap f) \multimap (t \multimap f) \qquad \dfrac{(k \multimap (Y \multimap X)) \multimap (Y \multimap X) \qquad \dfrac{\dfrac{\dfrac{[k]^1 \qquad k \multimap (g \multimap (t \multimap f))}{g \multimap (t \multimap f)}\multimap_E \qquad g}{t \multimap f}\multimap_E}{k \multimap (t \multimap f)}\multimap_{I,1}}{t \multimap f}\multimap_E}{t \multimap f}\multimap_E$$

$$\lambda s_0.\exists s[s \leq s_0 \wedge \forall x[girl(x, s) \rightarrow kiss(j, x, s)]]$$

Figure 2: *John sometimes kisses every girl* – Surface scope reading

$$\dfrac{(k \multimap (Y \multimap X)) \multimap (Y \multimap X) \qquad \dfrac{\dfrac{(t \multimap f) \multimap (t \multimap f) \qquad \dfrac{\dfrac{[k]^1 \qquad k \multimap (g \multimap (t \multimap f))}{g \multimap (t \multimap f)} \qquad g}{t \multimap f}\multimap_E}{t \multimap f}\multimap_E}{k \multimap (t \multimap f)}\multimap_{I,1}}{t \multimap f}\multimap_E$$

$$\lambda s_0.\forall x[girl(x, s_0) \rightarrow \exists s'[s' \leq s_0 \wedge kiss(j, x, s')]]$$

Figure 3: *John sometimes kisses every girl* – Inverse scope reading

2.4 Control and Raising verbs

Asudeh (2005) building on Landau (2003) illustrates how Glue semantic accounts for the differences between control and raising verbs ((11) and (12)). He compellingly argues that an analysis for control and raising comes naturally in Glue semantics, generalizing over the specific properties of both kinds of verbs. These specific properties arise from the fact that both of these constructions invoke structure sharing. In particular, the subject is shared between the matrix verb and the embedded verb.

(11) Gonzo tried to leave. control

(12) Gonzo seemed to leave. raising

The only difference between the two kinds of verbs lies in their PRED structure. Raising verbs, as the name suggest, treat the SUBJ as a thematic subject. This is the case for expletive or raising subjects (Asudeh 2005). On the other hand, the subjects of control verbs are thematic subjects as shown in Figure 4 on the right. How does this affect their semantics? We begin with explaining the semantic composition of raising verbs in the present framework.

$$\begin{bmatrix} \text{PRED} & \text{`seem<(XCOMP)>(SUBJ)'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED `Gonzo'} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`leave<(SUBJ)>'} \\ \text{SUBJ} & [\,] \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \text{PRED} & \text{`try<(SUBJ),(XCOMP)>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED `Gonzo'} \end{bmatrix} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`leave<(SUBJ)>'} \\ \text{SUBJ} & [\,] \end{bmatrix} \end{bmatrix}$$

Figure 4: Structure sharing in control and raising (Asudeh 2005)

The meaning constructor for our treatment of raising verbs, inspired by Asudeh (2005, 2002, 2000), is shown in (13). The crucial component is the linear logic side, which almost looks like a modifier, but it is anchored to two different TNS-ASP nodes. Note that the quantifier denotation proposed in the proof for adverbs now allows us to derive the scope ambiguity shown in (14). In fact, by comparison to control verbs, we will see that the meaning constructor presented above in (13) makes explicit the distinction between de dicto and de re interpretations in the semantics.

(13) $[\![\, \textit{seem} \,]\!] = \lambda p_{<s,t>}.\lambda s_0.\forall s'[s' \in SEEM(s_0) \rightarrow p(s')]$:
$(\uparrow \text{TNS-ASP}_{\text{XCOMP}} \multimap \uparrow \text{XCOMP}) \multimap (\uparrow \text{TNS-ASP}_{\text{f}} \multimap f)$
$= (i \multimap h) \multimap (t \multimap f)$

(14) Every girl seems to laugh.

a. every(girl,seem(laugh))
b. seem(every(girl(laugh)))

The proofs for the two readings are given in Figure 5 and 6. As illustrated above, the verb *seem* takes a proposition as its argument which, in combination with the NP quantifier, makes two readings available. Control verbs on the other hand do not allow for scope ambiguity with subject NP quantifiers. We contrast raising and control verbs by making raising verbs take propositions as their argument and control verbs verbs take properties as their argument. This is discussed next.

$$\dfrac{(g \multimap (Y \multimap X)) \multimap (Y \multimap X) \qquad \dfrac{\dfrac{\dfrac{[g]^1 \qquad g \multimap (i \multimap h)}{i \multimap h}\multimap_E \qquad (i \multimap h) \multimap (t \multimap f)}{t \multimap f}\multimap_E}{g \multimap (t \multimap f)}\multimap_{I,1}}{t \multimap f}\multimap_E$$

$$\lambda s_0.\forall x[girl(x, s_0) \rightarrow \forall s'[s' \in SEEM(s_0) \rightarrow laugh(x, s')]]$$

Figure 5: *Every girl seems to laugh* – Surface scope reading

$$\dfrac{(i \multimap h) \multimap (t \multimap f) \qquad \dfrac{(g \multimap (Y \multimap X)) \multimap (Y \multimap X) \qquad g \multimap (i \multimap h)}{i \multimap h}\multimap_E}{t \multimap f}\multimap_E$$

$$\lambda s_0.\forall s'[s' \in SEEM(s_0) \rightarrow \forall x[girl(x, s') \rightarrow laugh(x, s')]]$$

Figure 6: *Every girl seems to laugh* – Inverse scope reading

Comparing control and raising verbs in a situation semantics account has lead us an interesting puzzle with respect to the difference between the two. To understand this, let us look at our proposed denotation for control verbs:

(15) ⟦ *try* ⟧ = $\lambda P_{<e,<s,t>>}.\lambda x_e.\lambda s_0.\forall s'[s' \in INT_x(s_0) \rightarrow P(x)(s')]$:
(↑ XCOMP SUBJ ⊸ (↑ TNS-ASP$_{\text{XCOMP}}$ ⊸ ↑ XCOMP))
⊸ (↑ SUBJ ⊸ (↑ TNS-ASP$_{\text{f}}$ ⊸↑))
$= (g \multimap (i \multimap h)) \multimap (g \multimap (t \multimap f))$

Following Asudeh (2000, 2002, 2005), the control verb *try* is treated as taking a property of type $< e, < s, t >>$ as argument. The puzzle lies in the relation between the situation variable and the SUBJ that the antecedent and the consequent of the linear logic formula are bound to. This remains implicit in the work by Asudeh (2000, 2002, 2005). Concretely, in the previously cited work, properties could be understood as having a compound type in the consequent of their linear logic representation: $g_e \multimap (h_s \otimes f_t)$. Our own proposal, on the other hand, can be understood as strengthening the relation between the SUBJ and the situation variable, by illustrating how each element is bound in the f-structure.

However, note that we co-bind the individual type variable and the situation variable with the matrix verb. Thus, the result is a semantics of *try* whose complement is neither a property nor a proposition, but a truth value. While this point requires more research, we tentatively assume that this allows us to explain why raising complements allow for aspectual modification, while control verbs do not.[7]

[7]Following this idea, aspectual features would be modifiers from propositions to propositions. The following data can explained by this:

In this paper we pay particular attention to the scopal restrictions of control and raising verbs. As Asudeh (2005) notes "scopal elements can take both wide and narrow scope with respect to raising verbs but can only take wide scope with respect to control verbs". This observation is in agreement with (4) from Kallmeyer & Romero (2008) and is captured in the approach presented above. A proof using the control verb *try* is shown in Figure 7. In contrast to the previously discussed raising verb *seem*, only one reading is available as predicted.

$$\dfrac{(g \multimap (Y \multimap X)) \multimap (Y \multimap X) \qquad \dfrac{g \multimap (i \multimap h) \qquad (g \multimap (i \multimap h)) \multimap (g \multimap (t \multimap f))}{g \multimap (t \multimap f)} \multimap_E}{t \multimap f} \multimap_E$$

$$\lambda s.\forall x[girl(x,s) \rightarrow \forall s'[s' \in INT_x(s) \rightarrow laugh(x,s')]]$$

Figure 7: Glue proof without meanings: *Every girl tries to laugh.*

2.5 Attitude verbs

In contrast to control and raising verbs (in English), attitude verbs embed a finite clause. We already established that NP quantifiers are restricted to the first finite clause they occur in. However, we did not provide a concrete implementation for this. The intuitive idea is, that in LFG, the boundaries of such a clause are provided by the governing f-structure. In the XLE grammars of the ParGram project, finite clauses are distinguished by having a `CLAUSE-TYPE` feature. This does not exist in f-structures embedded under e.g., control and raising verbs, which, in English, describe non-finite clauses. Consequently, a quantificational NP can scope over a control verb like *try* and over a raising verb like *seem*, but not over an attitude verb like *think*. Thus, only the surface scope reading should be possible for (16).

(16) Mary thinks John loves every girl.

To model this difference between the verbs involving functional control and attitude verbs in the Glue semantics analysis of this paper, the lexical entry for the quantificational NP has to be modified in the scope of a `COMP`, if we want to preserve the general idea pursued in this paper. Thus, we require a mechanism that anchors the variables scoping over situations in the denotation of quantifiers to specific Glue constants, i.e. `TNS-ASP` nodes. This makes the situation variable the pivotal element in scope restrictions (section 3.1).

(i) a. John seems to have visited everybody. Kallmeyer & Romero (2008)
 b. Mary seems to be going to the park.

(ii) a. ??John tries to have visited everybody.
 b. #John tries to be going to the park.

$$
f\left[\begin{array}{lll}
\text{PRED} & \text{`think<Mary,love>'} & \\
\text{TNS-ASP} & t\left[\begin{array}{l}\text{TENSE pres}\\ \text{PROG -\_, PERF -\_}\end{array}\right] & \\
\text{SUBJ} & g\left[\text{PRED `Mary'}\right] & \\
\text{COMP} & k\left[\begin{array}{ll}
\text{PRED `love<girl,horse>'} & \\
\text{TNS-ASP} & j\left[\begin{array}{l}\text{TENSE pres}\\ \text{PROG -\_, PERF -\_}\end{array}\right]\\
\text{OBJ} & d\left[\begin{array}{ll}\text{PRED `horse'} & \\ \text{SPEC} & \left[\text{DET} \left[\text{PRED `a'}\right]\right]\end{array}\right]\\
\text{SUBJ} & e\left[\begin{array}{ll}\text{PRED `girl'} & \\ \text{SPEC} & \left[\text{QUANT} \left[\text{PRED `every'}\right]\right]\end{array}\right]
\end{array}\right] &
\end{array}\right]
$$

Figure 8: *Mary thinks every girl loves a horse.*

We illustrate the general idea in terms of example (17). It has two readings, as *a* can scope over *every* within the embedded finite clause. This means, within the scope boundary provided by the `COMP` quantifiers are to work as expected.

(17) Mary thinks every girl loves a horse.

(18) a. $[\![$ *think* $]\!] = \lambda p_{<s,t>}.\lambda x_e.\lambda s_0.\forall s'[s' \in DOX_x(s_0) \rightarrow p(s')]$:
$(\uparrow \text{TNS-ASP}_{\text{COMP}} \multimap \uparrow \text{COMP})$
$\multimap (\uparrow \text{SUBJ} \multimap (\uparrow \text{TNS-ASP}_{\text{f}} \multimap \uparrow))$
$= (j \multimap k) \multimap (g \multimap (t \multimap f))$

b. $[\![$ *every girl* $]\!] = \lambda Q_{<e,<s,t>>}.\lambda s.\forall x[girl(x,s) \rightarrow Q(x)(s)]$:
$\forall X.((\uparrow \text{COMP SUBJ} \multimap (\uparrow \text{TNS-ASP}_{\text{COMP}} \multimap X))$
$\multimap (\uparrow \text{TNS-ASP}_{\text{COMP}} \multimap X))$
$= \forall X.((e \multimap (j \multimap X)) \multimap (j \multimap X))$

c. $[\![$ *a horse* $]\!] = \lambda Q_{<e,<s,t>>}.\lambda s.\exists x[horse(x,s) \wedge Q(x)(s)]$:
$\forall X.((\uparrow \text{COMP OBJ} \multimap (\uparrow \text{TNS-ASP}_{\text{COMP}} \multimap X))$
$\multimap (\uparrow \text{TNS-ASP}_{\text{COMP}} \multimap X))$
$= \forall X.((d \multimap (j \multimap X)) \multimap (j \multimap X))$

The combination steps are those for the typically expected quantifier ambiguity within the complement (see Figure 9 and 10). The result is then simply combined with the attitude verb, which in turn is combined with the subject of the matrix sentence to yield the final result.

$$
\begin{prooftree}
\AxiomC{$(e \multimap (j \multimap k)) \multimap (j \multimap k)$}
\AxiomC{$(d \multimap (j \multimap X)) \multimap (j \multimap X)$}
\AxiomC{$[d]^1$}
\AxiomC{$d \multimap (e \multimap (j \multimap k))$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$e \multimap (j \multimap k)$}
\AxiomC{$[e]^2$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$j \multimap k$}
\RightLabel{$\multimap_{I,1}$}
\UnaryInfC{$d \multimap (j \multimap k)$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$j \multimap k$}
\RightLabel{$\multimap_{I,2}$}
\UnaryInfC{$e \multimap (j \multimap k)$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$j \multimap k$}
\AxiomC{$(j \multimap k) \multimap (g \multimap (t \multimap f))$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$g \multimap (t \multimap f)$}
\AxiomC{g}
\RightLabel{$\multimap_E$}
\BinaryInfC{$t \multimap f$}
\end{prooftree}
$$

every $>$ a:

$\lambda s_0.\forall s'[s' \in DOX_m(s_0) \rightarrow \forall x[girl(x, s') \rightarrow \exists y[horse(y, s') \wedge love(x, y, s')]]]$

Figure 9: Glue proof without meanings:
Mary thinks every girl loves a horse – Reading: *every* $>$ *a*

$$
\begin{prooftree}
\AxiomC{$(d \multimap (j \multimap k)) \multimap (j \multimap k)$}
\AxiomC{$[d]^1$}
\AxiomC{$d \multimap (e \multimap (j \multimap k))$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$e \multimap (j \multimap k)$}
\AxiomC{$(e \multimap (j \multimap X)) \multimap (j \multimap X)$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$j \multimap k$}
\RightLabel{$\multimap_{I,1}$}
\UnaryInfC{$d \multimap (j \multimap k)$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$j \multimap k$}
\AxiomC{$(j \multimap k) \multimap (g \multimap (t \multimap f))$}
\RightLabel{$\multimap_E$}
\BinaryInfC{$g \multimap (t \multimap f)$}
\AxiomC{g}
\RightLabel{$\multimap_E$}
\BinaryInfC{$t \multimap f$}
\end{prooftree}
$$

a $>$ every:

$\lambda s_0.\forall s'[s' \in DOX_m(s_0) \rightarrow \exists y[horse(y, s') \wedge \forall x[girl(x, s') \rightarrow love(x, y, s')]]]$

Figure 10: Glue proof without meanings:
Mary thinks every girl loves a horse – Reading: *a* $>$ *every*

3 GSWB Case Study

In this section we test the analysis presented above within the GSWB. We discuss the challenges that arose during this case study and present a solution based on Lev (2007). The case study follows a simple procedure. We designed a small Treebank and produced a list of premise sets for each sentence manually. Testing them in the workbench revealed a number of issues, the most crucial of which and their solution we discuss in the next section.

3.1 Scope restrictions

As discussed in the previous section, most of the scope predictions discussed in the previous section fall out naturally due to the chosen Glue semantics representation. However, the final example of attitude verbs does not fit into the general picture of this paper in the sense that we modified the involved NP quantifiers. In particular, we anchored the situation variable used in NP quantifiers to constants corresponding to certain `TNS-ASP` nodes.

We assume that this operation is done by means of some rewriting system (in particular, the rewrite system in XLE). As Gotham (2019) points out, this is not desirable from a theoretical point of view, but we want to avoid changing the linear logic fragment of the GSWB. This rules out the Gotham (2019) approach and leaves us with constraining either proofs or meaning constructors. We chose to do the latter. Concretely, quantifiers that occur with in a specific `COMP` are rewritten such that they are anchored to the specific `COMP`'s `TNS-ASP` node.

3.2 Typed semantics in the GSWB

This section describes the main results of our exploration of the capabilities of the workbench. The primary issue was that the GSWB could not deal properly with typed semantic representations. To understand this, let us first describe the implementation of the Hepple (1996) algorithm in the Glue semantics workbench Meßmer & Zymla (2018).

The algorithm presented in Meßmer & Zymla (2018) is a recursive algorithm following the proposal made by Hepple (1996) quite strictly. Higher-order premises (i.e. premises which have a linear implication as antecedent) are compiled. The compilation process is a simplification of complex linear logic formulas with the result of only having to deal with simple combination steps between atoms and corresponding linear implications. The main issue with the algorithm used in Meßmer & Zymla (2018) lies within this compilation process.

The process works as follows: the antecedent of a given complex premise is divided into antecedent and conclusion. This antecedent is cut off from the original formula. The result is two formulas: the remains of the original formula and a compiled-out assumption. The process is shown in (19).[8] The new assumption

[8]The indices in brackets are used to track premises. Introducing a new assumption also introduces

receives an unused variable as meaning representation. This variable is indirectly anchored to the original meaning representation. First, the whole representation is wrapped by a lambda binder which binds another unused variable. Second, the original meaning representation receives as argument a lambda function which binds the variable of the compiled out assumption and scopes over the previously introduced fresh variable.

(19) $\alpha : (a \multimap b) \multimap c$ [0] $\Rightarrow_{compile} \lambda u.\alpha(\lambda v.u) : b[a] \multimap c$ [0];
$v : \{a\}$ [1]

A formally undesirable consequence of Hepple's (1996) algorithm it, that it relies on accidental binding. The newly introduced lambda binder in the argument of the meaning representation α needs to bind the compiled-out variable v.[9] A concrete example is given in figure 11.

$$\dfrac{H[g_2] \multimap H : \lambda\text{u}.\lambda\text{P}.\forall\text{x}[\text{person(x)} \wedge \text{P(x)}](\lambda\text{v.u}) \qquad \dfrac{g_1 \multimap f : \lambda\text{y.sleep(y)} \qquad \{g_2\} : \text{v}}{f\{g_2\} : \text{sleep(v)}}}{\dfrac{f : \lambda\text{P}.\forall\text{x}[\text{person(x)} \wedge \text{P(x)}](\lambda\text{v.sleep(v)})}{f : \forall\text{x}[\text{person(x)} \wedge \text{sleep(x)}]}\ \beta\text{-conversion}}\ [\text{H/f}]$$

Figure 11: Every person sleeps. – Hepple style

This process works for compiling simple antecedents, however, the implementation in Meßmer & Zymla (2018) does not deal properly with the recursive nature of the algorithm. To illustrate this, consider the nominal quantifier *every girl* translated into a version that includes situations as semantic objects of type s in the ontology.

(20) a. $[\![$ *Every* $]\!] = \lambda P_{<e,<s,t>>}.\lambda Q_{<e,<s,t>>}\lambda s_s.\forall x[P(x)(s) \rightarrow Q(x)(s)]$
b. $[\![$ *student* $]\!] = \lambda x_e.\lambda s_s.student(x, s)$
c. $[\![$ *Every student* $]\!] = \lambda Q_{<e,<s,t>>}.\lambda s_s.\forall x[student(x, s) \rightarrow Q(x)(s)]$

This corresponds to the following linear logic formulas:[10]

(21) a. $g \multimap (h \multimap i)$
b. $(g \multimap (h \multimap i)) \multimap \forall X, Y.(j \multimap (t \multimap X) \multimap (t \multimap X))$

This requires a compilation step that is not anticipated in the algorithm that was previously implemented in the workbench. It is not difficult to imagine, how the compilation goes on the linear logic side: First the resource g would be compiled out. In a next step the resource h would be compiled out. This would result in two accidental lambda bindings which are in fact desired (Hepple 1996). However, the part of the linear logic formula that is quantified over also requires compilation.

a new index. However, this is not relevant for discussing the flaw of the algorithm.

[9] Thus, the functional application used within the linear lambda calculus (the lambda calculus used for the compilation) is different from classical functional application in so far that it does allow for this accidental binding. As a result of this, Hepple's (1996) original version is not readily compatible with out-of-the-box beta conversion tools. Another issue, that the present paper remedies.

[10] The issue occurs in particular in `COMP` embedded NP quantifiers, since our original quantifier denotation results in a modifier resource, which is not compiled (Meßmer & Zymla 2018).

This adds up to a total of four lambda binders, introducing four lambda functions in its scope. This means the compiled meaning representation (MR) of the quantifier in (20-a) is bound by four lambdas after the compilation steps.

(22) a. $\lambda m.\lambda p.\lambda r.\lambda u.MR(\lambda v.u)(\lambda s.r)(\lambda q.p)(\lambda n.m)(\alpha)$
b. $MR(\lambda v.\lambda s.\lambda q.\lambda n.\alpha)$

(23) a. $\lambda m_t.\lambda p_{s,t}.\lambda r_t.\lambda u_{s,t}.MR(\lambda v_e.u)(\lambda s_s.r)(\lambda q_e.p)(\lambda n_s.m)(\alpha)(\beta)$
b. $MR(\lambda n_s.\lambda q_e.\alpha)(\lambda v_e.\lambda s_s.\beta)$

The issue here is the number of arguments that are created during the compilation process. At first glance, (22) suggests, that there are four arguments to the denotation of the quantifier. However, it is apparent that there should only be two arguments (corresponding to λP and λQ in (20-a)). In the original single-type version of the prover, saturating the outermost lambda results in a number of undesired lambda applications. The result of these applications is, that there is only one argument, instead of two, that has been passed along through all the lambda slots. There are several possible solutions to this problem. We opt for a solution along the lines of Lev (2007).

However, first some fundamental issues have to be fixed. The first step is to properly type the lambda binders as in (23) to achieve the desired solution given in the example. This approach runs into a number of problems still. While the formula given in (23) technically works, deriving it via the compilation process is not straightforward. To solve this issue, let us look at it from another perspective:

There is a second problem with the original Hepple (1996) approach, namely, that it requires accidental binding. Lev (2007) presents a way to circumvent the need for this semantically "unsound" approach. The gist of his improved algorithm is that the meaning side of the compiled resource is not modified, but as before, the compilation process creates a new meaning side variable of the type of the resource that has been compiled out on the meaning side. However, this step does not coincide with adding a new argument to the meaning representation as done in (22) (the four lamba binders that apply to the meaning representation *MR* are introduced by the compilation step).

The role of making sure that the compiled out variables are inserted into the proof appropriately is taken by the discharge system in Lev (2007). Discharges mark where a specific variable has been compiled out. Now, if a variable that has been compiled out (or any element that has combined with that variable) combines with the meaning representation that has discharged it, a modified functional application procedure is applied. Lev (2007) formalizes this as shown in Figure 12.

$$\frac{\phi : A : S_1 \qquad \delta : A_L \rightarrow B : S_2}{\delta(\lambda v_{i_1}, ..., \lambda v_{i_n}.\phi) : B : S_1 \cup S_2} \quad \begin{array}{l}\text{provided } S_1 \cap S_2 = \varnothing \text{ and } L \subset S_1 \\ \text{and } L = [i_1, ..., i_n]\end{array}$$

Figure 12: Functional Application in Lev (2007) style compilation

Given a function and an argument with compatible linear logic resources, the two are combined only if the discharges of the function are a subset of the assumptions made for the argument. In that case, for all discharges a lambda introduction step is executed. This requires the discharges to be a list rather than set, to preserve appropriate types. Elements that are cut off (i.e. discharged) have to be reintroduced in the reverse order. The whole procedure is exemplified in (24) to (26).

(24) **Premises:** $a_e \multimap b_t \multimap f_t : \alpha$
$a_e \multimap b_t : \beta$

(25) **Compiled:** $b_t[a] \multimap f_t : \alpha$
$\{a_e\} : v_e$
$a_e \multimap b_t : \beta$

(26)
$$\dfrac{\dfrac{a_e \multimap b_t : \beta \qquad \{a_e\} : v_e}{b_t\{a_e\} : \beta(v_e)} \qquad b_t[a] \multimap f_t : \alpha}{f_t : \alpha(\lambda v_e.\beta(v_e))}$$

3.3 Further additions to the GSWB

This section very briefly describes some additional features of the workbench which were added when implementing the verbal scoping analysis described in this paper. These involve a debugging mode and improved file handling.

The debugging mode provides some basic information about performance of the system. This is illustrated in Figure 13. As shown there, the computation time is measured. Additionally, information about attempted inference (combination) steps is collected. This includes those leading up to the proof as well as those that are not used in the final derivation. Furthermore, the number of compilation steps is counted. This metric has been added to provide transparency about the conversion from higher-order linear logic formulas to first-order formulas.

```
Debugging report:
The following data was collected:
computationTime: 12ms
Number of iterations through Sequent: 18
Number of combination steps: 12
Number of proper compilation steps:2
```

Figure 13: Debugging mode sample output

In addition to the debug mode, several file handling features have been added. Primarily, the system now allows the user to specify input and output files, which makes it easier to use the workbench in a pipeline architecture. Furthermore, the workbench now also allows the user to process multiple proofs from a single file, another functionality that has been developed with the creation of pipelines in mind. In general, the modular nature of the GSWB was been improved.

4 Conclusion

In this paper we presented a case study of computational Glue semantics that serves as the first use case of the Glue semantics workbench. This study was based on two-type situation semantics and allowed us to cover phenomena in the domain of quantification over and within certain kinds of verbs, such as control and raising verbs. Our main goal was to explore scope interactions between quantificational NPs and kinds of verbs that are semantically quantifiers. We have shown that many scope interactions can be derived by anchoring quantifiers to specific `TNS-ASP` nodes, which are assumed to map to situation variables. We highlighted further possible roles of `TNS-ASP` nodes in the distinction of control vs. raising verbs. Furthermore, we established that quantifier scope cannot be simply restricted by `TNS-ASP` nodes when embedded under an attitude verbs. This is due to the fact that quantifiers do not distinguish between matrix clause NPs and `COMP` clause NPs. We have provided a tentative technical solution but more work needs to be done in this area.

By virtue of implementing a situation semantics approach for Glue, we could test the GSWB intensively. Although the case study has shown that there is a lot of work yet to be done with respect to the GSWB, it helped get a better understanding of what has to be done. The result is a more functionally robust and flexible system for working with Glue semantics. The next important step is to improve the syntax/semantics interface. Both co-descriptive and description-by-analysis approaches are currently being developed and help to push the improvement of the GSWB.

References

Asudeh, Ash. 2000. Functional Identity and Resource-Sensitivity in Control. In *Proceedings of the LFG00 Conference*, 2–24.

Asudeh, Ash. 2002. A Resource-Sensitive Semantics for Equi and Raising. In David Beaver, Stefan Kaufmann, Brady Clark & Luis D. Casillas Martínez (eds.), *The Construction of Meaning*, 1–21. Stanford, CA: CSLI Publications.

Asudeh, Ash. 2005. Control and Semantic Resource Sensitivity. *Journal of Linguistics* 41(3). 465–511.

Barwise, Jon. 1981. Scenes and Other Situations. *The Journal of Philosophy* 78(7). 369–397.

Bennett, Michael & Barbara Hall Partee. 1978. Toward the Logic of Tense and Aspect in English. In Barbara H Partee (ed.), *Compositionality in Formal Semantics*, vol. 84, Blackwell Publishing Ltd.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The Parallel Grammar Project. In *Proceedings of the 2002 Work-*

shop on Grammar Engineering and Evaluation, vol. 15, 1–7. Association for Computational Linguistics.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press on Demand.

Crouch, Richard & Josef van Genabith. 2000. Linear Logic for Linguists. *URL: http://www2. parc. com/istl/members/crouch* .

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34. New York: Academic Press.

Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1999. Quantification, Anaphora, and Intensionality. In Mary Dalrymple (ed.), *Semantics and Syntax in Lexical Functional Grammar – The Resource Logic Approach*, 39–89.

Gotham, Matthew. 2019. Constraining Scope Ambiguity in LFG+Glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'19 Conference*, CSLI Publications.

Haug, Dag. 2008. Tense and Aspect for Glue Semantics: The Case of Participial XADJ's. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, 291–311. Cent. Study Lang. Inf. Stanford, CA.

Heim, Irene. 2011. Definiteness and Indefiniteness. In Paul Portner, Klaus von Heusinger & Claudia Maienborn (eds.), *Semantics – Noun Phrases and Verb Phrases*, .

Hepple, Mark. 1996. A Compilation-Chart Method for Linear Categorial Deduction. In *Proceedings of the 16th Conference on Computational Linguistics-Volume 1*, 537–542. Association for Computational Linguistics.

Hintikka, Jaakko. 1969. Semantics for Propositional Attitudes. In Donald Davidson, Jaako Hintikka, G. Nuchelmas & Wesley C. Salmon (eds.), *Models for Modalities*, 87–111. Dordrecht: D. Reidel Publishing Company.

Kallmeyer, Laura & Maribel Romero. 2008. Scope and Situation Binding in LTAG Using Semantic Unification. *Research on Language and Computation* 6(1). 3–52.

Kratzer, Angelika. 2019. Situations in Natural Langugae Semantics. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2019 edn. `https://plato.stanford.edu/archives/sum2019/entries/situations-semantics/`.

Landau, Idan. 2003. Movement out of Control. *Linguistic Inquiry* 34(3). 471–498.

Lev, Iddo. 2007. *Packed Computation of Exact Meaning Representations*: Stanford University dissertation.

Lowe, John. 2014. Gluing Meanings and Semantic Structures. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, CSLI Publications.

Meßmer, Moritz & Mark-Matthias Zymla. 2018. The Glue Semantics Workbench: A Modular Toolkit for Exploring Linear Logic and Glue Semantics. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'18 Conference, University of Vienna*, 249–263. Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/LFG/2018/lfg2018-messmer-zymla.pdf.

Part II
Contributions to the Workshop

Achievement Predicates and Tense Paradigms in Hazaragi

Saira Bano
University of Konstanz

Miriam Butt
University of Konstanz

Ashwini Deo
Ohio State University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 394–414

http://csli-publications.stanford.edu/LFG/2019

Keywords: Hazaragi, tense/aspect, Achievements, vowel system, underspecification, morphology-syntax, lexical semantics

Bano, Saira, Butt, Miriam, & Deo, Ashwini. 2019. Achievement Predicates and Tense Paradigms in Hazaragi. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 394–414. Stanford, CA: CSLI Publications.

Abstract

In this paper we investigate and model an interaction between lexical semantics and morphological tense paradigms in Hazaragi, an under-described Eastern Iranian language that is closely related to Dari. Hazaragi is spoken in Afghanistan and Pakistan, but also world-wide among the Hazara diaspora. Hazaragi is an SOV language with a complex tense/aspect paradigm. The paradigm exhibits a split in that it shows a past/non-past distinction with certain verbs, while with other verbs the non-past morphology is more restricted. We offer a semantic explanation for the observed pattern and show that achievement verbs lie at the heart of the split. We formally model this interaction between morphology, syntax and lexical semantics within LFG's modular architecture of grammar. In this, we rely on the proposal for the morphology-syntax interface developed by Dalrymple (2015) and the system for crosslinguistic annotation and calculation of tense/aspect developed by Zymla (2018). We implement a ParGram style grammar fragment for Hazaragi (Crouch et al. 2017, Butt et al. 1999) to model the interface between morphology, syntax and lexical semantics.

1 Introduction

Hazaragi is an Eastern Iranian language spoken mainly in the Central Afghanistan region known as Hazarajat (Dulling, 1973) and in the city of Quetta in Southern Pakistan.[1] Due to the Hazara diaspora, the language is also found across the world. Hazaragi is very close to Dari, one of the national languages of Afghanistan. It also shares features with Kabuli (Farhadi 1975), the contact vernacular in the country. While Kieffer (2003) reports Hazaragi as a dialect, *Ethnologue Languages of the World* recognizes Hazaragi as a language with a population of 2,295,000 speakers. The ethnic group is known as Hazaras or Hazareh, but the group refers to themselves as Azra and to their language as Azergi.

Hazaragi displays a complex tense/aspect system. It has two verbal stems, the present and the past stem. In this paper, we focus on a split whereby a past/non-past distinction is found in one part of the verbal paradigm, but not in another part. In example (1a), we see that the present tense form of the verb *eat* can be used to convey both ongoing present (habitual/generic or event-in-progress) temporal reference and future temporal reference. In contrast, as seen in (1b) and (1c), the past stem is restricted to past temporal reference. The imperfective can appear on both present and past stems, but it does not lead to non-past readings with the past stem. Thus, the difference in stems effects a past/non-past distinction.[2]

(1) a. ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

[1] We gratefully acknowledge support for this research from the DAAD, the German Academic Exchange Office.

[2] The abbreviations in the glosses are as follows: Impf=Imperfective, Imp=Imperative, Subj=Subjunctive, Nom=Nominative, Pres=Present, Sg=Singular, DOM=Direct object marker.

b. ali nan mu-xord
Ali food Impf-eat.Past.3Sg
'Ali was eating food.'

c. ali nan xord
Ali food eat.Past.3Sg
'Ali ate food.'

However, the paradigm plays out differently with a certain group of verbs, as shown in (2). Here the present tense form of the verb cannot be used to convey an event-in-progress reading. It can only have habitual/generic or future reference.

(2) a. ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali comes home (habitually)/#Ali is coming home/Ali will come home.'

b. ali xana amad
Ali house come.Past.3Sg
'Ali came home.'

The paradigms for the two verbs in our examples, namely *xordan* 'to eat' and *amadan* 'to come', are shown in Table 1. As can be seen, exactly the same morphology, namely the -*d*+Pers/Num affixes vs. an *mV*- prefix give rise to different tense/aspect distinctions, depending on which verb they occur with.

	xor 'eat'		*a* 'come'	
Pers/Num	**Past**	**Non-Past**	**Past**	**Non-Pres**
1st Sg.	xordʊm	mʊxrʊm	amadʊm	majʊm
2nd Sg.	xordi	mʊxri	amadi	maji
3rd Sg.	xord	mʊxra	amad	maja
1st Pl.	xordi	mʊxri	amadi	maji
2nd Pl.	xordɪn	mʊxrɪn	amadɪn	majɪn
3rd Pl.	xordən	mʊxrən	amadən	majən

Table 1: Different Tense Distinctions within the Verbal Paradigm

In section 2 we show that this split in the readings follows from an interaction of tense/aspect meanings with the Aktionsart of the verb. Achievement verbs pattern as in (2), while all other verb classes pattern as in (1). We provide a semantic explanation for this: achievements, which mostly describe punctual, non-durative eventualities, typically lack a process sub-event. When combined with present tense morphology, such verbs do not describe an eventuality in progress at reference time, since such meanings presuppose that the eventuality referred to is durative. It is this contrast between the temporal contours associated with achievement predicates vs. durative predicates (activities, accomplishments, and states) that is responsible for the effect. While at first glance, this may appear to be a distinctly different system, one observes that achievement predicates in English inflected with present progressive marking often seem to resist a standard event-in-progress reading as well. Bach (1986) distinguishes between "culminations" and "happenings" within the class of achievements. Culminations are those predicates that describe achievements that have an associated process (*reach, die*) while happenings are 'lucky' achievements in which there is no associated process that leads to the transition described by the predicate (*notice, recognize, flash*). With culminations, progressive marking in the present tense typically induces a future-oriented reading — the event is understood to obtain in the future. With happenings, there is the added possibility of an iterative reading in which the predicate describes a sequence of repeated events. The upshot is that the peculiar temporal contour of achievement predicates is partly responsible for the unavailability of the standard event-in-progress reading associated with present marking with durative predicates. This is systematically evident in Hazaragi.

In analyzing this distinction in the verbal paradigm of Hazaragi, we were faced with complex morphophonology. We provide a first pass at an analysis of the Hazaragi verbal morphophonology in section 3.

In section 4 we show how our analysis can be implemented in a ParGram style grammar fragment for Hazaragi (Crouch et al. 2017, Butt et al. 1999). In order to model the interface between morphophonology, syntax and semantics, we rely on the proposal for the morphology-syntax interface developed by Dalrymple (2015). In section 5, we show how the tense/aspect distinctions could be calculated via Zymla's (2018) multi-tiered analysis of tense.

2 The Asymmetry in Temporal Reference

In this section we provide necessary background information on the structure of Hazaragi grammar with respect to the verbal system.

2.1 The Hazaragi Verbal Paradigm

Hazaragi is an SOV language, with both prefixes and affixes attaching to a verbal stem. The verbal paradigms are constructed on the basis of two stems: the present and the past stem. The past stem is realized via affixation of /-ɪd/ to the base form. The allomorphy in this domain is seen in Table 2. The regular verbs in lines 1–3 surface with the allomorph /-ɪd/. For irregular verbs as in the lines 4–5 the affix surfaces as /-d/ after /n/ and /r/, while after /s,ʃ,z,f/, it surfaces as /-t/ as can be seen with /ʃ/ in lines 6–7. Lines 7–8 give examples of some irregular verbs where the pattern is unpredictable.[3]

[3]There are several takes on the relationship of the stems to one another (see Farahani (1990) for an overview). One analysis postulates that the two different stems for regular verbs derive from two separate sources, another assumes that the past stem is derived from the form of the present stem via

Verb stems are further inflected for person and number. There is no gender distinction in the language. The suffixes for person and number for the past and the present are the same except for the third person singular, which has an overt inflection for the present stem, but not for the past stem (see Table 1).

	Present Stem	**Past stem**	**Gloss**
1.	bar	barɪd	pour
2.	xar	xarɪd	buy
3.	poʃ	poʃɪd	wear
4.	xor	xord	eat
5.	kan	kand	pluck
6.	koʃ	kuʃt	kill
7.	el	eʃt	keep
8.	g <gu	guft	say
9.	ʃn[4]<ʃaw	ʃud	happen

Table 2: Sample Verbs with their Present and Past Stems

There are three verbal prefixes in Hazaragi: the imperfective, the imperative/subjunctive and a marker of negation. Negation is expressed via *na*, but we are not concerned with it here. The imperfective and imperative/subjunctive surface in a variety of forms. The initial consonant is invariable: /m/ for the imperfective and /b/ for the imperative/subjunctive. This consonant is followed by a vowel which shows great variance, but this variance appears to be phonologically conditioned. We therefore analyze these two prefixes as having an underlyingly underspecified vowel: /mV-/ and /bV-/.

Imperfective marking appears on both present and past stems and gives rise to habitual/generic and progressive readings. The subjunctive and imperative are constructed by attaching the prefix /bV-/ to the present stem. While the subjunctive has forms across all the person-number cells, imperative forms are only available for the 2nd person singular and plural.

We propose that the underspecified vowel in both the imperfective /mV-/ and the imperative/subjunctive /bV-/ is subject to assimilation with the first vowel contained in the stem, thus being realized variably as [m-], [ma-], [me-], [mɪ-], [mʊ-] and [mo-]. We provide the details and motivation for this analysis in section 3, but first turn to explaining the tense split.

---

affixation (e.g., Lazard 1992), while another analysis (Henderson 1978) posits one underlying root which is form-identical to the non-past stem. The past stem is then taken to be derived via the affixation of an archisegment /D/, which is realized variously as /-d,-t,-ɪd/ depending on the phonological environment. The irregular verbs have no predictable pattern. Cowen and Yarmohammadi (1978) formulate three rules to derive the past stem from the underlying root.

[4]The deletion of /aw/ and insertion of /n/ indicates that several phonological processes have applied. We are currently studying these.

2.2 Explaining the Asymmetry in Temporal Reference

As already mentioned in the introduction, Hazaragi exhibits a past/non-past distinction for most verbs, as seen in (3) (repeated from above). The present tense form is ambiguous with respect to temporal reference and disambiguation can be effected via overt temporal expressions, as shown in (4).

(3) a. ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

b. ali nan xord
Ali food eat.Past.3Sg
'Ali ate food.'

(4) a. ali darau nan mʊ-xr-a
Ali now food Impf-eat.Pres-3Sg
'Ali is eating food now.'

b. ali saba nan mʊ-xr-a
Ali tomorrow food Impf-eat.Pres-3Sg
'Ali will eat food tomorrow.'

In contrast, as shown above in (2) and repeated here in (5), with some verbs, the present tense form is not compatible with an event-in-progress reading. It only yields future-oriented or habitual/generic readings, as shown in (6).

(5) a. ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali comes home (habitually)/Ali will come home/#Ali is coming home.'

b. ali xana amad
Ali house come.Past.3Sg
'Ali came home'

(6) ali ameʃa ʃau m-aj-a
Ali always night Impf-come.Pres-3Sg
'Ali always comes in the evening.'

2.2.1 Durative vs. Non-durative Verbal Predicates

A closer look at the class of verbs that fail to give rise to the event-in-progress reading with present morphology shows that it is exactly those verbs which denote achievements in the sense of Vendler (1957). Achievement verbs are taken to describe punctual events, specifically a transition to some result state. Although there is a sense in which an accompanying process is presupposed in the use of such verbs, such a process is not taken to be lexicalized in the verb's meaning. What is critical here is that achievement predicates, in contrast to other aspectual classes, denote punctual eventualities that do not unfold over an interval of time. A useful diagnostic for identifying achievement predicates is the interaction between these and temporal expressions such as *for an hour* and *in an hour*. As has been noted starting with Dowty (1979), achievement predicates are unacceptable with *for an hour* type of temporal expressions, which presuppose atelicity. With *in an hour* type of temporal expressions, they give rise to a reading distinct from accomplishments, conveying that the eventuality described by the verb occurred *after* one hour from some salient reference time in the past. We see for Hazaragi that the achievement verb 'reach' is unacceptable with the durative *for an hour* and gives rise to the 'after an hour' reading with the *in an hour*.

(7) a. ʈrain da station məne jak genʈa rasıd
train at station in one hour reach.Past.3Sg
'The train reached the station in an hour(after an hour).'

b. *ʈrain da station bleje jak genʈa rasıd
train at station for one hour reach.Past.3Sg
'The train reached the station for an hour.'

Almost the same pattern of interpretation and acceptability is exhibited for *amadan* 'to come', shown in (8). The difference is that while the collocation with *for an hour* is unacceptable on the reading that the coming eventuality lasted an hour, it is acceptable on the reading that the result-state of the coming eventuality (Ali's being home) lasted for an hour. Piñon (1999) has observed that this is an available reading for durative expressions with achievement predicates. In contrast, *xordan* 'to eat' in (9) is an activity predicate (unless combined with a quantized object argument). It is acceptable with the durative *for an hour* and conveys that the eating eventuality went on for an hour. It is less acceptable with *in an hour*.

(8) a. ali məne jak genʈa xana amad
Ali.Nom in one hour house come.Past.3Sg
'Ali came home in an hour.'

b. ali bleje jak genʈa xana amad
Ali for one hour house come.Past.3Sg
'Ali came home for an hour (he stayed for an hour).'

(9) a. *ali məne jak genʈa nan xord
Ali in one hour nan eat.Past.3Sg
'Ali ate food in an hour.'

b. ali bleje jak genʈa nan xord
Ali for one hour nan eat.Past.3Sg
'Ali ate food for an hour.'

Other stative/activity predicates such as *ʃɪʃtan* ‘to sit’ and *malidan* ‘to rub’, which are clearly durative and atelic, pattern like *xordan* ’to eat’, see (10) and (11).

(10) a. ∗ali məne jak genʈa ʃɪʃt
Ali in one hour sit.Past.3Sg
‘Ali sat in an hour.’

b. ali bleje jak genʈa ʃɪʃt
Ali for one hour sit.Past.3Sg
‘Ali sat for an hour.’

(11) a. ∗ali məne jak genʈa malam xo-r malɪd
Ali in one hour ointment his-DOM rub.Past.3Sg
‘Ali rubbed his ointment in an hour.’

b. ali bleje jak genʈa malam xo-r malɪd
Ali for one hour ointment his-DOM rub.Past.3Sg
‘Ali rubbed his ointment for an hour.’

In contrast, accomplishment predicates, illustrated in (12) by *dʒersi baftan* ‘to knit a jersey’ denote telic events that have both a process and a result component. As seen in (12), such a predicate is acceptable with *in an hour* but unlike with achievements, the sentence conveys that the event of knitting a jersey was accomplished within that amount of time. The combination with *for an hour* is slightly less acceptable but not ungrammatical.

(12) a. maaham məne jak genʈa dʒersi baft
Maaham in one hour jersey knit.Past.3Sg
‘Maaham knitted a jersey in an hour.’

b. #maaham bleje jak genʈa dʒersi baft
Maaham for one hour jersey knit.Past.3Sg
‘Maaham knitted a jersey for an hour.’

Further verbs which pattern like *amadam* ‘to come’ (i.e., as achievements) are: *eʃtan* ‘to keep’, *awurdan* ‘to bring’, *rasɪdan* ‘to reach’, *rasandan* ‘to cause to reach’, *xastan* ‘to ask’, *poʃɪdan* ‘to wear’, *zadan* ‘to hit’, *murdan* ‘to die’ *kuʃtan* ‘to kill’. The number of simple verbs in Persian numbers in the hundreds (rather than the thousands) and we expect a similar situation in Hazaragi, given the trouble we had identifying verbal predicates that were not complex predicates.

2.2.2 Further Evidence from Complex Predicates

This fundamental pattern in the language also holds for complex predications, as illustrated in (13) and (14) for two different N-V complex predicates. In (13) with the light verb ‘hit’, the present event-in-progress reading is unavailable. In (14) with the light verb ‘do’, in contrast, the reading becomes available.

(13) Ali bini mɪ-zn-a
Ali nose Impf-hit.Pres-3Sg
'Ali will sneeze./#Ali is sneezing.'

(14) Ali darga ra taxtax mʊ-n-a
Ali door DOM knock Impf-do.Pres-3Sg
'Ali will knock at the door./Ali is knocking at the door.'

Light verbs have been shown to determine the overall Aktionsart of the complex predication (Butt 1995) and this is the case in Hazaragi as well. As illustrated in (15), when the light verb is an achievement predicate (*zadan* 'to hit'), the complex predicate is also an achievement, as shown by the interpretations associated with the temporal expressions. In (15a), we get the 'after an hour' reading. It is relevant to note here that for a class of achievement predicates, the combination with *for an hour* gives rise to the iterative reading rather than unacceptability.[5] This is the case with (15b), which conveys that there was repeated sneezing over the course of an hour.

(15) a. ali məne jak genʈa bini zad
Ali in one hour nose hit.Past.3sg
'Ali sneezed in an hour (after an hour).'

b. ali bleje jak genʈa bini zad
Ali for one hour nose hit.Past.3sg
'Ali sneezed for an hour (kept sneezing over and over).'

But when the light verb denotes an activity, e.g., *kadan* 'to do' as in (17), the complex predicate patterns with all non-achievement verbs.

(16) a. ∗ali məne jak genʈa darga ra taxtax kad
Ali in one hour door DOM knock do.Past.3sg
'Ali knocked at the door in an hour.'

b. ali bleje jak genʈa darga ra taxtax kad
Ali for one hour door DOM knock do.Past.3sg
'Ali knocked at the door for an hour.'

Having understood the reason for the different tense readings in the Hazaragi verbal paradigm, we now first turn to investigating the complex morphophonology (section 3) and then provide an implementation that models the interaction between lexical semantics, tense interpretation, and morphophonology in section 4.

[5] Such predicates denote punctual events but do not encode a transition to a result state. These are known as semelfactives (Comrie 1976, Smith 1991) or happenings (Bach 1986).

3 Hazaragi Verbal Morphophonology

The morphophonological patterns within the Hazaragi verbal paradigm are complex. Recall that the imperfective prefix can be attached to either the present or past verbal stem and that this in turn is inflected for person and number. This results in the following pattern: m+vowel+stem+suffix. The imperative and subjunctive are both expressed via a b- prefix. This b- prefix exhibits exactly the same morphophonological patterns as the imperfective prefix: b+vowel+stem+suffix. An overview of the patterns for the third person singular with the present and the past stems is provided in Tables 3 and 4.

	Past Stem	**Impf. 3rd Sg**	**Transl.**
1.	bʊrd	mʊ- bʊrd	take
2.	poʃɪd	mo-pʃɪd	wear
3.	amad	me-mad	come
4.	eʃt	me-ʃt	keep
5.	xord	mʊ- xord	eat
6.	raft	mo-raft	go

Table 3: Imperfective Forms with Past Stem for the 3rd Singular

The forms in Table 4 illustrate the variation found in the surface realization of the imperfective prefix with the past stem. This variation can be accounted for through a process in which the underspecified vowel of the prefix m+vowel acquires the place features of the stem vowel as in lines 1, 2 and 4. But in lines 3, 5 and 6, we find that the prefix holds a different vowel then the stem vowel. It is seen that the stem vowels are deleted in lines 2–4.

We further investigated the present imperfective where the variation in the prefix vowel is found at its most. Some sample verbs illustrating the variation are shown in Table 5.

	Present Stem	**Impf. 3rd Sg**	**Transl.**
1.	bər<bur	mʊ- bra	take
2.	pəʃ<poʃ	mo-pʃa	wear
3.	a <ɑj	ma-ja	come
4.	el	me-la	keep
5.	xər<xor	mʊ- xra	eat
6.	r<row	mo-ra	go

Table 4: Imperfective Forms with Present Stem for the 3rd Singular

The table illustrates several complications. One concerns the identification of the underlying vowel in the present stem. In several cases this has been lost and/or a schwa surfaces for purposes of adhering to syllabic phonotactic constraints. We therefore adduced evidence from New Persian and Middle Persian verbal stems (as well as the corresponding Hazaragi past stems) to establish the underlying forms, indicated by the $<$ in the tables. After reconstructing the Hazaragi present stem it is evident that the underlying vowel emerges as the prefix vowel and as the stress has to be on the prefix, the stem vowel deletes from the verbal stem as in lines 3, 4 and 6. Here resyllabification also takes place and the stem consonant becomes the coda of the first syllable as in lines 1, 2 and 5.[6]

As shown in Table 5, another complication is the patterns found with the stem vowel /a/. Instead of /a/ surfacing as the prefix vowel, several different variants in terms of mɪ-, me-, mʊ- and mo- are possible.

	Present Stem	**Impf. 3rd Sg**	**Transl.**
7.	xər<xar	me-xra	buy
8.	məl <mal	mʊ- mla	rub
9.	tərs <tars	mɪ- tərsa	scare
10.	ʃan<neʃɑn	mɪ- ʃana	sit (caus)
11.	pər <par	mo-pra	‘fell‘

Table 5: Stems containing an /a/ Vowel with the Imperfective

In order to understand this variation, we investigated the Hazaragi vowel system in some detail.

3.1 The Hazaragi Vowel System

The Hazaragi vowel chart in Figure 1 was constructed with original data elicited from Hazaragi speakers.[7] It constitutes the first phonological analysis of the Hazaragi vowel system.[8]

There are 8 vowels in Hazaragi i, ɪ, e, ə, a, u, ʊ, o and 24 consonants.[9] We see the vowels as being divided via a primary three-way distinction of height where all the vowels above the (dotted line) are high (i, u) and high-mid and the vowels below are low (ə, a). The most interesting are the high mid vowels /e, ɪ, ʊ, o/ as they are very close to another in height. We also see a further division into front (i, ɪ, e) and back vowels (u, ʊ, o). The /ə/ is central but at a low mid position while /a/ is low but not completely at the back or front. We therefore analyze /a/ as placeless.

We adopt the *Featurally Underspecified Lexicon* (FUL) model (Lahiri and Evers 1991, Lahiri and Reetz 2002, Scharinger et al. 2010). FUL analyzes consonants and vowels via the same place features. The place features are divided into two further sub-nodes: articulator and tongue height. This means that the height features [high] and [low] can be specified independently of the place of articulation,

[6]This is true for the bisyllabic words. For the trisyllabic words the stem vowel deletion is not seen and we are currently further investigating the stress pattern and resyllabification patterns of Hazaragi.

[7]These were average of 600 tokens of 40 monosyllabic words each repeated 5 times and elicited from three speakers (one male, two female). The vowel chart is constructed on scatter plot where the F1 (formants for tongue height) was graphed on the y axis and F2 (formants for tongue frontness) on the x axis and rotated afterwards so to give the exact location of the vowel in the oral cavity.

[8]At this point we would like to thank Aditi Lahiri for providing essential guidance in pointing us towards the potentially relevant factors in the Hazaragi vowel system.

[9]We have not added the consonant chart due to space limitations.

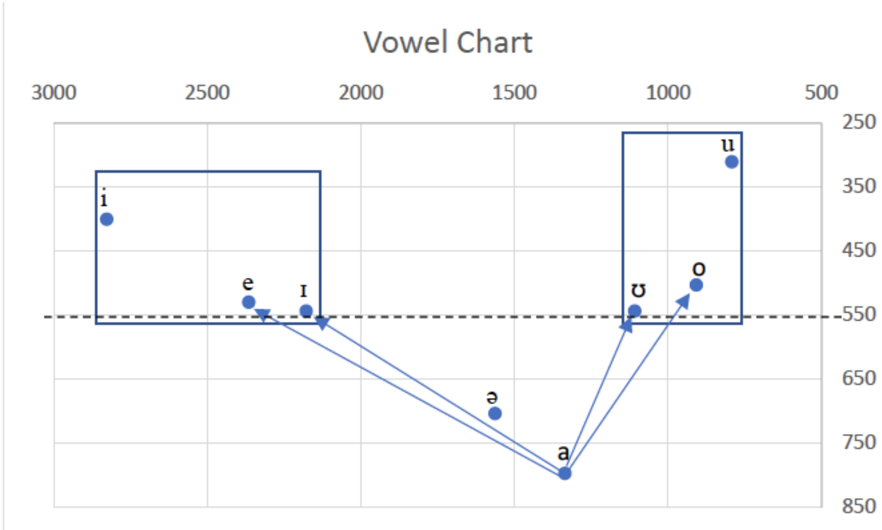


Figure 1: Hazaragi Vowel Chart

e.g. coronal, labial and dorsal. FUL also states that the height features will both be specified if the language has a three-way height distinction (Lahiri and Reetz 2010). The feature *mid* is only used for descriptive purposes. The front vowels are analyzed as coronal, the back vowels as dorsal, while rounded vowels are labial.

Applying this model to Hazaragi, we take the high-mid vowels to be underspecified for height and the low /ə, a/ vowels to be underspecified for place of articulation. Note that the high-mid vowels are the ones which emerge as the specification of the prefix vowel, which we take to be underspecified for place of articulation. Lahiri and Reetz (2002, 2010) postulate that coronal is an underspecified feature and Ghini (2001) further argues that a segment which is not specified for a place of articulation is underlyingly a coronal by default. Adopting Ghini's analysis for Hazaragi means that we assign the feature coronal to /a/ and our underspecified prefix vowel mV- by default.

An analysis based on FUL suggests separating out the surface variants of mV- into two groups of coronal vs. labial/dorsal. The [mɪ-] and [me-] are coronal, while the [mʊ-] and [mo-] are labial/dorsal vowels. The forms in Table 3 and 4 fall out naturally under this analysis: when the stem vowel is labial/dorsal, the prefix vowel is labial/dorsal. When the stem vowel is coronal, the prefix vowel is coronal. Recall that when /a/ appears as the stem vowel, the mV- surfaces variously as mɪ-, me-, mʊ- and mo-. Our analysis as the coronal being specified by default accounts for the coronal mɪ- and me- variants in Table 5.

However, something extra must be said for the mʊ- and mo- variants. With respect to these we posit that when a stem labial consonant follows the mV- it provides an environment along with the [m] of the prefix for the formation of a labial vowel in the prefix: mʊ- and mo- as in *mʊmla* in line 8 of Table 5 and the examples in Table 6.[10]

	Present Stem	**Impf. 3rd Sg**	**Transl.**
12.	xər <xor	mʊ- xra	eat
13.	xər <xar	me-xra	buy
14.	pər <par	mo-pra	fell

Table 6: Imperfectives with Variant Forms

This initial analysis for the morphophonology of the Hazaragi verbal system suffices as a basis for this paper; In the next section, we present an implementation that models the complex interaction between lexical semantics, tense interpretation and the morphophonology of the verbal forms.

4 The Morphology-Syntax Interface

We implement our analysis via the XLE grammar development platform (Crouch et al. 2017) and integrate a finite-state morphological analyzer for Hazaragi that is based on the design and implementation of finite state morphological (FSM) analysis described in Beesley and Karttunen (2003). While we present our analysis in computational terms, we note that our treatment of Hazaragi follows the formal theoretical model for an integration of morphology into LFG as laid out by Dalrymple (2015). In that model the morphological component is assumed to specify a morphological realization relation *R*, which associates a Lexemic Index, a s(emantic)-form, and a p(honological)-form with a set of m(orphological)-features. The Lexemic index is simply an arbitrary label that is used to identify the lexeme. For our two example Hazaragi verbs *xordan* 'to eat' and *amadan* 'to come' in the imperfective present form, the morphological component relates the information in (17) to one another via the *R* relation. For the past form, the information in (18) is placed in correspondence to one another.

(17) a. M-entry for the word form *mʊxra*:
R<XOR, eats, /mʊxra/,
{M-CAT:VERB,M-STEM:PRES,M-ASP:IMPF,M-PERS:3,M-NUM:SG}

b. M-entry for the word form *maja*:
R<A, comes, /maja/,
{M-CAT:VERB,M-STEM:PRES,M-ASP:IMPF,M-PERS:3,M-NUM:SG}

(18) a. M-entry for the word form *xord*:
R<XOR, ate, /xord/,
{M-CAT:VERB,M-STEM:PAST,M-PERS:3,M-NUM:SG}

[10]/x/ is a velar in Hazaragi, but as shown that only the labial vowels or labial stem initial consonant can provide the mʊ- or mo- prefix.

b. M-entry for the word form *amad*:
R<A, came, /amad/,
{M-CAT:VERB,M-STEM:PAST,M-PERS:3,M-NUM:SG}

In terms of the actual FSM analysis, the finite state machine places the overt word form in correspondence with the morphological analysis as shown below. The past forms are relatively straightforward. The forms with the imperfective prefix require a more complex treatment.

```
xord:xor+Verb+Past+3P+Sg
amad:a+Verb+Past+3P+Sg
muxra:xor+Verb+Pres+Impf+3P+Sg
maja:a+Verb+Pres+Impf+3P+Sg
```

Within the FSM, the imperfective prefix is analyzed as an /m/ with an underspecified vowel: mV. The FSM specifies the list of vowels available in Hazaragi and further subdivides these into coronal (ɪ,e) and labial vowels (ʊ, o) as per the analysis in section 3. The surface realization of this underspecified vowel is calculated via a set of phonological rules that are sensitive to the first vowel and the initial consonant (labiality) found in the underlying root form of the verb.

The information from the FSM is integrated into the Hazaragi grammar via the Description function *D*, which maps a set of m-features to the appropriate c-structure category and f-descriptions. In the *dogs* example of Dalrymple (2015), this looks as in (19): For the Lexemic index DOG1 that the m-features M-CAT and M-NUM are associated with, the appropriate c-structure category is N and the f-structural information is that the number is plural. This information is passed into the grammar as part of the morphology-syntax interface.

(19) *D*<DOG1,{M-CAT:NOUN,M-NUM:PL},N,{(↑NUM)=pl}>

With respect to our examples, the Description function *D* relates the information shown in Table 8 to the morphological tags (m-features) coming out of the FSM. In the XLE implementation, the Description function is realized in terms of lexical entries at the sublexical level. That is, the morphological tags are treated as lexical entries with which c-structural (category) and f-structural information can be associated. For example, the entry for +Past below says that this is of the category TENSE and is associated with the f-structural information [TENSE past]. The 'xle' tells the grammar that this is information coming out of the morphological analzyer. For a more detailed description of the computational realization of this morphology-syntax interface, see Kaplan et al. (2004).

The morphological tags +Verb and +Noun do not have any f-structural information associated with them: in this analysis they simply provide category information. The +Impf tag, on the other hand, carries a complex f-structural specification. It contains a disjunction (signaled via the { | }) that says that either a future reading or a present reading is possible, but that the present tense reading is only possible if the verb does not denote an achievement. In addition, a past tense reading is possible if licensed by the use of the past stem.

+Verb	V-T	xle	
+Noun	N-T	xle	
+1P	V-PERS	xle	(↑SUBJ PERS) = 1
+2P	V-PERS	xle	(↑SUBJ PERS) = 2
+3P	V-PERS	xle	(↑SUBJ PERS) = 3
+Sg	V-NUM	xle	(↑SUBJ NUM) = sg
+Pl	V-NUM	xle	(↑SUBJ NUM) = pl
+Past	V-STM	xle	(↑TNS-ASP TENSE) = past
+Pres	V-STM	xle	(↑TNS-ASP TENSE) $\neq$ past
+Impf	ASP	xle	{ (↑TNS-ASP TENSE) = pres
			(↑TNS-ASP AKTIONSART) $\neq$ achievement
			| (↑TNS-ASP TENSE) = fut
			| (↑TNS-ASP TENSE) =c past }

Table 7: The Description function *D*

The items specified in the sublexical lexicon need to be parsed so that only well-formed sequences of tags are allowed. This is achieved via sublexical rules that formally work just like phrase structure rules (Kaplan et al. 2004). The sublexical rule for the Hazaragi verb in our analysis is shown in Table 8. This rule expects a verb to consist of a verb stem (V-S), a verb tag (V-T), information about whether it is the present or past form of the verb stem (V-STM), an optional specification for aspect (ASP; e.g., the mV- prefix), and then person and number information.[11]

V	$\longrightarrow$	V-S_BASE	"verb stem"
		V-T_BASE	"category verb"
		V-STM_BASE	"past or present stem"
		(ASP_BASE)	"optional aspect (impf)"
		V-PERS_BASE	"person"
		V-NUM_BASE.	"number"

Table 8: Sublexical Rules

The verb stems themselves are specified in the lexicon as shown in Table 8. The lexical entries pick up on the lemma (Lexemic index), the category (V-S), the fact that this information is coming out of the FSM (xle) and associate the lemma

[11]The addition of _BASE is required for XLE internal reasons in the implementation to identify these rules as sublexical. We have kept them in the example so as not to sow confusion if grammar writing with XLE is attempted along the lines described in this paper. A reviewer also notes that one could in principle collapse the stem, aspect and person/number tags into one tag (V-TAG, for example) and let the FSM ensure the correct order and type of the tags. However, our experience with Urdu grammar development showed that a detailed articulation of the sublexical rules as in Table 8 did serve to constrain the grammar further.

and the category information with f-structural descriptions that include information about the predicate-argument structure and the Aktionsart of the verb.

Note that this information about the lexical semantics of a verb cannot be extracted from the morphological analyzer (since there is no morphological marking of valency or Aktionsart in Hazaragi), but should be added as part of the verbal lexicon in the grammar.

xor	V-S	xle	(↑PRED) = 'xor<(↑SUBJ)(↑OBJ)>' "eat" (↑TNS-ASP AKTIONSART) = activity.
a	V-S	xle	(↑PRED) = 'a<(↑SUBJ)(↑OBL)>' "come" (↑TNS-ASP AKTIONSART) = achievement.

Table 9: Lexical Entries for Verb Stems

With the information coming from the morphological analyzer and the information contained in the verbal stem lexicon, all the information is in place in order to provide the right analysis for the asymmetry in available readings in the Hazaragi verbal paradigm. The phrase structure rules needed for the analyses of our example sentences are simple and straightforward (and also similar to those of the Urdu ParGram grammar, see Butt and King (2007), so we do not show them here).

The f-structures in Figures 2 and 3 respectively show the analyses for the main example sentences, repeated here in (20) and (21).

(20) ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food/ Ali is eating food/ Ali will eat food.'

(21) ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali will come home/#Ali is coming home.'

As can be seen, the combination of the imperfective prefix mV- and a non-achievement verb like 'eat' leads to ambiguity — the sentence can be interpreted either as denoting an event in the present or in the future. With an achievement verb like 'come', on the other hand, only the future reading is obtained.

This section has shown how a complex interplay between morphophonology, lexical semantics and tense readings can be modeled via a conception of the morphology-syntax interface as set out in Dalrymple (2015). For the sake of concreteness and systematic testing, we have implemented this analysis via the XLE grammar development platform, adhering to standards developed as part of the ParGram effort (Butt et al. 1999).

We have, however, so far had nothing to say about the generic/habitual readings that can still be expressed, even with achievement verbs. In the next section, we introduce Zymla's multi-tiered analysis of tense/aspect and show how this reading can be calculated on the basis of morpho-syntactic information.

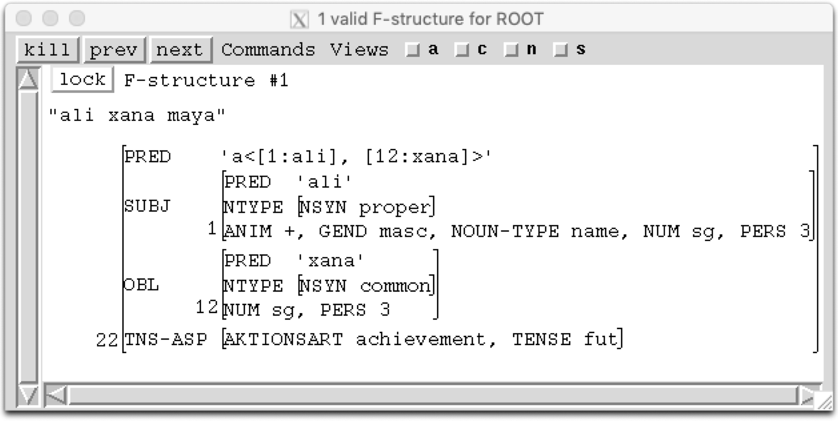


Figure 2: Present Event-in-Progress Reading Not Available

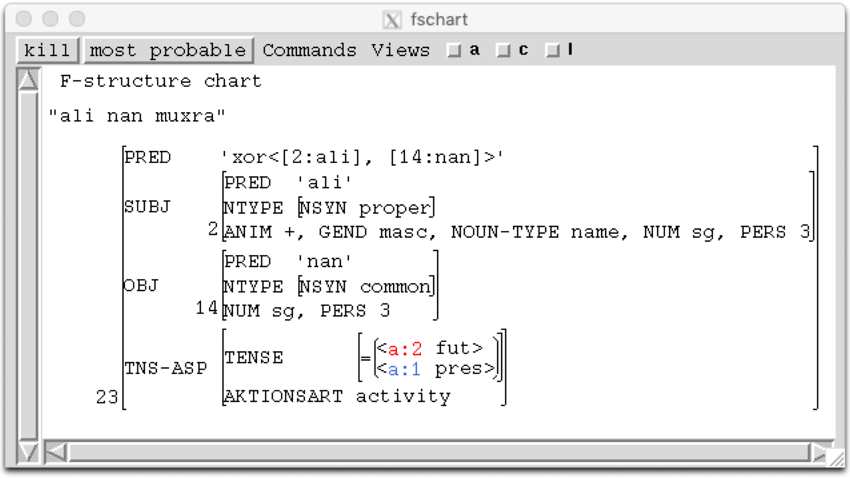


Figure 3: Present Event-in-Progress Reading Available

5 Zymla's Multi-Tiered Analysis of Tense/Aspect

5.1 The Basic System

Zymla (2018) is concerned with providing a crosslinguistically and computationally viable annotation scheme for the calculation of tense/aspect semantics. His proposal is compatible with LFG and, in particular, with ParGram style grammars (Zymla and Sulger 2017). Zymla proposes three tiers in his scheme: Tier 1 picks up on the overtly available morphosyntactic cues in a clause. For example, the verbal morphology in a sentence like *John left* indicates a simple past tense, which is registered at f-structure as [TENSE past]. On the basis of this f-structural information the semantic information can be calculated that the time of the event is to be interpreted as having occurred before the current time: $\lambda t.t \prec t_0$. A more complex situation arises when the information relevant for a calculation of the tense/aspect semantics is derived from several different parts of the morphosyntax and is combined to derive a new meaning. An example provided by Zymla concerns habitual readings as in *John builds houses.* In this case, there is a [TENSE pres] and a plural object. From these two pieces of morphosyntactically encoded information it can be concluded that the reading is one of a habitual imperfective. This more complex calculation via a set of implication rules is accomplished at Tier 2 of Zymla's system. Tier 3 provides the locus for further pragmatic reasoning that involves putting together the information gleaned from the analysis levels provided by Tiers 1 and 2 with further information found in the clause or in the context. Zymla provides the German example in (22) as an illustration.

(22)	John	komm-t	morgen	an.
	John.Nom	come-Pres.3Sg	tomorrow	on
	'John arrives tomorrow.'			

As in Hazaragi, the German present morphology is in principle compatible with present event-in-progress, future and habitual/generic readings. However, unlike Hazaragi, it does not exhibit an asymmetry in readings available. In (22) the cue for disambiguation is provided by the temporal adverbial *morgen* 'tomorrow', which provides a temporal reference point that is located in the future. This part of the reasoning about the temporal information is located at Tier 3 in Zymla's system.

This multitiered system has the advantage of providing a way of calculating temporal semantic knowledge only on the basis of simple overt morphosyntactic cues (Tier 1), a combination of cues (Tier 2) and the integration of further contextual information that triggers a pragmatic reasoning component (Tier 3). If further contextual information is not available, then only reasoning up to Tier 2 is necessary. If there is no complex information to be combined, then Tier 1 calculations are sufficient. The system thus provides a nice level of "back-off" strategies depending on what type of morphosyntactic and contextual information is available.

5.2 Application to Hazaragi

In the analysis presented in section 4, the morphology-syntax interface provided f-structural tense information and already factored in the disambiguation to only future readings when the verb is an achievement (Figure 2). The analysis in section 4 had nothing to say about the habitual/generic readings that are also possible across all verb classes.

Zymla's system provides us with a slightly different way forward. Recall our central examples, repeated here in (23) and (24). An analysis at Tier 1 yields the result that we have [ASPECT imperfective] due to the mV- prefix and that we have [TENSE pres] due to the present form of the stem. A Tier 2 analysis will yield the information that the interpretation could be either in the present or in the future or be a progressive or habitual/generic predication. It will also exclude the present and progressive interpretation for achievement verbs, as this information can be factored in by implication rules at Tier 2 that pick up on information encoded at f-structure ([AKTIONSART achievement]) via the lexical semantic specifications.

(23) ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

(24) ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali will come home/#Ali is coming home.'

Finally, calculations at the level of Tier 3 could disambiguate the possible readings by taking further contextual information such as the presence of temporal expressions as in (25) and (26) into account (repeated from above).

(25) ali ameʃa ʃau m-aj-a
Ali always night Impf-come.Pres-3Sg
'Ali always comes in the evening.'

(26) a. ali darau nan mʊ-xr-a
Ali now food Impf-eat.Pres-3Sg
'Ali is eating food now.'

b. ali saba nan mʊ-xr-a
Ali tomorrow food Impf-eat.Pres-3Sg
'Ali will eat food tomorrow.'

That is, these Tier 2 and Tier 3 calculations could be performed on the basis of the overtly available morphosyntactic information from the clause outside of the f-structural analysis. The f-structure would only represent information at the level of Tier 1 and leave the more complex calculations at Tier 2 and Tier 3 for a semantic and pragmatic component that bases itself on f-structure information, but also goes beyond it as proposed, for example, in the Abstract Knowledge Representation (AKR) approach to semantics proposed in Bobrow et al. (2007).

6 Conclusion

In this paper we have investigated the verbal paradigm of Hazaragi. We identified a split in the availability of tense readings in one and the same morphological paradigm and showed that this could be explained by taking the Aktionsart of the verbs into account. Achievement verbs lack a durative component and thus do not allow for a present tense event-in-progress reading, whereas all other verbs do.

In studying the verbal paradigm, we were confronted with complex morphophonology. The adoption of the FUL model for the Hazaragi vowel system led us to postulate an underlyingly underspecified vowel in the imperfective prefix: mV-. We further assume that /a/ in Hazaragi is a placeless vowel and that vowels are specified as coronal by default. We argued that the underspecified vowel in the prefix acquires place features from the verbal stem vowel as well as the initial stem consonant when that is a labial. The prefix vowel thus variably surfaces as mɪ-, me-, mʊ- and mo-, as conditioned by the phonology of the stem.

With both the semantic insight and the morphophonological analysis in place, we then went on to show how the complex interaction between morphophonology, lexical semantics and tense/aspect semantics can be modeled via the morphosyntax interface as defined by Dalrymple (2015) and how tense interpretation can proceed via the multi-tier system proposed by Zymla (2018).

References

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 1, 5–16.

Beesley, Kenneth and Karttunen, Lauri. 2003. *Finite-State Morphology*. Stanford: CSLI Publications.

Bobrow, Daniel G., Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's Bridge and Question Answering System. In Tracy Holloway King and Emily M. Bender (eds.), *Proceedings of the GEAF07 Workshop*, pages 46–66, Stanford: CSLI Publications.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam and King, Tracy H. 2007. Urdu in a Parallel Grammar Development Environment. In T. Takenobu and C.-R. Huang (eds.), *Language Resources and Evaluation*, volume 41, pages 191–207, Special Issue on Asian Language Processing: State of the Art Resources and Processing.

Butt, Miriam, King, Tracy Holloway, Niño, Maria-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

Cowen, Romayne and Yarmohammadi, Lotfollah. 1978. The Persian Verb Reconsidered. *Archiv Orientalni* 46, 46–60.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy Holloway, Maxwell III, John T. and Newman, Paula. 2017. XLE Documentation, https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle_toc.html.

Dalrymple, Mary. 2015. Morphology in the LFG Architecture. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG'15 Conference*, pages 64–83, Stanford: CSLI Publications.

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Farahani, Ali Akbar Khomeijani. 1990. *A Syntactic and Semantic Study of the Tense and Aspect System of Modern Persian*. Ph. D.thesis, The University of Leeds.

Farhadi, A. G. 1975. *The Spoken Dari of Afghanistan: A Grammar of Kāboli Dari (Persian), Compared to the Literary Language*. Kabul.

Ghini, Mirco. 2001. Asymmetries in the phonology of Miogliola. In *Studies in Generative Grammar 60*, Berlin: Mouton de Gruyter.

Henderson, Michael M. T. 1978. Modern Persian Verb Morphology. *Journal of the American Oriental Society* 98, 375–388.

Kaplan, Ronald M., III, John T. Maxwell, King, Tracy Holloway and Crouch, Dick. 2004. Integrating Finite-state Technology with Deep LFG Grammars. In *Proceedings of the European Summer School on Logic, Language and Information (ESSLLI) Workshop on Combining Shallow and Deep Processing for NLP*.

Kieffer, Charles M. 2003. HAZĀRA iv. Hazāragi dialect. *Encyclopaedia Iranica* XII, 90–93.

Lahiri, A. and Evers, V. 1991. Palatalization and coronality. In C. Paradis and J.-F. Prunet (eds.), *The Special Status of Coronals*, pages 79—-100, San Diego: Academic Press.

Lahiri, Aditi and Reetz, Henning. 2002. Underspecified recognition. *LabPhon* 7, 637–676.

Lahiri, Aditi and Reetz, Henning. 2010. Distinctive Features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38, 44–59.

Lazard, Gilbert. 1992. *A Grammar of Contemporary Persian*. Costa Mesa: Mazda Publishers.

Piñon, Chris. 1999. Durative Adverbials for Result States. In S. Bird, A. Cake, J. Haugen and P. Norquest (eds.), *WCCFL 18 Proceedings*, pages 420–433, Cascadilla Press.

Scharinger, Mathias, Lahiri, Aditi and Eulitz, Carsten. 2010. Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics* 23(4), 1–17.

Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66(2), 143–160.

Zymla, Mark-Matthias. 2018. Annotation of the Syntax/Semantics interface as a Bridge between Deep Linguistic Parsing and TimeML. In *Proceedings of the 14th Joint ACL–ISO Workshop on Interoperable Semantic Annotation*, pages 53–59, Association for Computational Linguistics.

Zymla, Mark-Matthias and Sulger, Sebastian. 2017. Cross-Linguistically Viable Treatment of Tense and Aspect in Parallel Grammar Development. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'17 Conference*, pages 406–426, Stanford: CSLI Publications.

The Mismatch between Morphological Symmetricality and Syntactic Ergativity in Pazeh

Li-Chen Yeh
The Australian National University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 415–430

http://csli-publications.stanford.edu/LFG/2019

Keywords: Austronesian voice system, argument structure, symmetricality, ergativity

Yeh, Li-Chen. 2019. The Mismatch between Morphological Symmetricality and Syntactic Ergativity in Pazeh. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 415–430. Stanford, CA: CSLI Publications.

Abstract

This paper discusses the interaction and the mismatch between a syntactically ergative system and morphologically symmetrical markings in Kulon-Pazeh (Austronesian language, ISO: UUN, Taiwan, henceforth Pazeh), specifically referring to the effect of voice morphology on the surface realisation of arguments. Based on the evidence from Pazeh, I argue that there is no one-to-one correspondence of symmetricality or correlation between syntax and morphology. At the morphology level, Pazeh demonstrates symmetricality for the markings on the verbs in actor voice, undergoer voice and instrumental voice, but at the syntactic level, the surface realisation of arguments shows that the system is, in fact, ergative, and hence deeply asymmetrical. This study shows that the operation of the Pazeh voice system contains a mismatch of symmetricality between syntax and morphology, involving co-present parallel structures which can be best illustrated using the Lexical Functional Grammar (LFG) framework. The finding contributes to the empirical understanding on voice alternations and the alignment system of Austronesian languages of Taiwan.

1 Introduction[1]

Voice systems are a typical feature of many Austronesian languages. For Philippine-type voice systems, it is commonly claimed that the semantic role of the subject is indicated by the affix on verb (Himmelmann, 2005; Zeitoun & Huang, 1997). The notion of symmetricality of Austronesian voices has been discussed jointly at both the syntactic level and the morphological level in the literature. The definition varies from author to author. According to the definition given by Himmelmann (2005: 113–114), having a symmetrical voice system may refer to having at least two voice alternations with morphological markings on the verb, and neither of the verb forms is a derived form of another. Symmetricality at the syntactic level is also observed in some Western Austronesian languages, also known as the Indonesian-type, where none of the voices has their nominal arguments taking overt marking by prepositions or case markers (Arka, 2003; Himmelmann, 2002: 11 , 2005: 112). In Balinese, actor and patient arguments of transitive verbs can be equally selected as the grammatical subject without demotion of the other argument (Arka, 2019). However, unlike the Indonesian-type languages, Pazeh is

[1] I would like to thank the anonymous external and internal reviewers for their comments which contribute to the improvement of this paper. All remaining errors are my own.

categorised as the Philippine-type. Based on the above definition of symmetricality, Pazeh only demonstrates symmetricality for voice alternations at the morphological level, not at the syntactic level. That is, the morphosyntactic operations for voices in Pazeh reveal a mismatch between morphological symmetricality and syntactic ergativity. Traditional terminology used by previous studies on this language has been hindering the search for a suitable morphosyntactic category for its voice system, and often led to confusion with the type of alignment system whether it is symmetrical, accusative or ergative.

This paper is organised as follows: In section 2, I present an overview of the morphological marking for three identified voices in Pazeh, illustrating the symmetrical markings at the morphological level. In section 3 and 4, I provide an overview of the phrasal markers and grammatical functions in Pazeh, showing evidence for a syntactically ergative system in Pazeh. In section 5, I provide empirical data for pronominals to illustrate how the pronominal paradigm in Pazeh supports an ergative analysis. Concluding remarks are given in section 6.

2 Pazeh voices and morphological affixation

Pazeh is an indigenous language of Taiwan, which used to be spoken in the northwest plains. It can be considered as a language with a null pronoun (i.e. pro-drop languages discussed in (Falk, 2006: 49–60)). Previous studies on Pazeh including Lin (2000) and Li and Tsuchida (2001, 2002) have identified at least three voices in Pazeh[2] (i.e. *mV-*, *-en*, *sa(a)-…*(*-an*)/ *si-…-*(*-an*)). The affixes are used to denote actor voice, undergoer voice and instrumental voice respectively. To illustrate, in actor voice, the affix selects a nominal phrase to be the subject whose semantic role is *Agent*, as in (1a) and the example sentence (1d). Patient voice affix signals the subject whose semantic role is *Patient*, as in (2b) and the sentence (2d), whereas instrumental voice affix indicates the subject whose semantic role is *Instrument*, as in (3c) and the example (3d).

[2] Lin (2000) and Li and Tsuchida (2001, 2002) applied the “focus system” framework, containing terminology specifically developed for describing the Austronesian languages of the Philippines. In their analysis, Pazeh demonstrates four types of verbal construction for voices, namely, agent focus, patient focus, instrumental focus and locative focus. Please refer to Ross and Teng (2005) for clarification on the differences between common terminology and those used by Formosan linguists. Due to insufficient empirical data to justify the status of locative voice, I will only discuss the morphosyntactic operation for actor voice, undergoer voice and instrumental voice in this paper.

(1) -baket ‘hit’

a.	mu-baket	b.	baked-en	c.	saa-baket
	‘AV.hit’		‘UV.hit’		‘IV.hit’

d.	mu-baket	(a)	rakihan	ki	aba
	AV-hit	OBL	child	ABS	father.DEF

‘The father beat a child.’

(Li & Tsuchida, 2001: 81)

(2) -xe’et ‘tie (with.something)’

a.	me-xe’et	b.	xe’ed-en	c.	saa-xe’et
	‘AV.tie’		‘UV.tie’		‘IV.tie’

d.	xe’ed-en	ni	Awi	ki	wazu
	tie-UV	ERG	person.name	ABS	dog.DEF

‘Awi put a leash on the dog.’

(Li & Tsuchida, 2001: 322)

(3) -te’eng ‘throw’

a.	me-te’eng	b.	te’eng-en	c.	si-te’eng
	‘AV.throw’		‘UV.throw’		‘IV.throw’

d.	si-te’eng	(a)	wazu	ni	rakihan	ki	batu
	IV-throw	OBL	dog.INDEF	ERG	child.DEF	ABS	stone.DEF

‘The child throw the stone at a dog.’

(Li & Tsuchida, 2001: 32)

A survey of verb roots with high semantic transitivity shows that the morphological forms of Pazeh verb are shown to be equally marked for these three voices. Li and Tsuchida (2001) provided rich evidence to support this claim, as cited in example (4) to (7) below.

(4) -ken/-kan 'eat'

a.	me-ken	b.	kan-en	c.	saa-ken-an
	'AV. eat'		'UV.eat'		'IV.eat'

(5) -kixis/-kexes 'cut'

a.	mu-kixis	b.	kexez-en	c.	saa-kixis
	'AV. cut'		'UV.cut'		'IV.cut'

(6) -kizu 'dig'

a.	mu-kizu	b.	kizu'-un	c.	saa-kizu
	'AV.dig'		'UV.dig'		'IV.dig'

(7) -talek 'cook'

a.	mu-talek	b.	talek-en	c.	saa-talek
	'AV.cook'		'UV.cook'		'IV.cook'

An investigation on the effect of voice affixation shows that different morphological markings on the verb reflect different selection of the arguments to be the pivot. The notion of pivot has been applied to the discussion of the Philippine-type voice systems by Foley and Van Valin (1984). In the case of Tagalog, the selection of an argument of a verb in the voice system (a.k.a. focused NP) is considered to involve pragmatic factors (e.g. definiteness) within a clause. Therefore, the selected argument is considered to show a functional similarity to English topics, where its discourse status and the syntactic status are operated under what Foley and Van Valin (1984: 115) called the "pragmatic pivots." This notion of pivot is introduced into LFG by Manning (1994, 1996) under the discussion of the inverse mapping theory, and later expanded by Falk (2006) to distinguish the properties of pivots from those of argumenthood. Falk's theory of pivot highlights the adaptability of the pivothood to account for languages that are not argument-pivot languages (e.g. Mandarin), where pivothood seems unrelated to argument mapping (Falk, 2006: 206). For the purpose of discussing the flexibility of turning certain arguments into pivot at the a-structure, as in the case of Pazeh, I adopt the idea of "a-subject" proposed by Manning (1996) and "l-subject (i.e. logical-subject)" mentioned in Arka (2003) in this paper. Manning (1994, 1996) differentiated the a-subject from the grammatical subject (i.e. g-subject or surface GF-SUBJ). It is considered in this paper that these different concepts of subjects are helpful to reveal the interaction between morphology and syntax with different subjects at different structures (*a-str*, *f-str* and *sem-str*).

In the case of Pazeh voice alternation, an actor argument is selected as the grammatical subject or pivot in actor voice, whereas in non-actor voices, a patient/ an instrument is selected. However, further investigation in the following sections shows that in addition to the grammatical subject or pivot selection triggered by the voice morphology on the verb, these equally marked morphological affixations will also trigger both promotion and demotion of an argument. The distribution of phrasal markers serves as a clue to reveal the non-symmetrical pattern of syntax in Pazeh.

3 Phrasal markers

The use of phrasal markers as a preliminary reference to the syntactic status of arguments is not uncommon in the literature of Austronesian languages in Taiwan. Often these phrasal markers are known as case markers as in Li and Tsuchida (2001: 32) and Lin (2000: 123). According to Li and Tsuchida (2001, 2002), Pazeh allows all kinds of phrasal markers to be omitted in natural speech, but when the markers are present, they have different forms, indicating different cases. Therefore, the absence of phrasal markers does not mean that the case marking is not in effect because these markers are usually presented in careful speech. Pronominal forms also support this claim, which will be discussed in section 5. However, even though both Li and Tsuchida (2001, 2002) and Lin (2000) observed the patterns of phrasal markers, the syntactic functions of the arguments involved in voice alternations remain unexplored in the existing literature.

Among all the phrasal markers shown in Table 1, Li and Tsuchida (2001, 2002) use the terms: nominative case for the argument selected by the voice morphology, genitive case for the agent in undergoer voice, oblique case for the patient in actor voice and locative case for locative argument, but they did not explicitly express the grammatical functions encoded by each case marker, and the type of voice alignment under their analysis remains unknown.

Table 1. Li and Tsuchida (2001: 32–33): Pazeh phrasal markers

Nominative	Genitive	Oblique	Locative
ki	ni	u/a	di

Lin (2000) diverges from Li and Tsuchida (2001, 2002) in that the patient argument in actor voice is labelled as accusative case, and she specifically analysed the argument denoted by nominative case as the subject, and the one marked by accusative case marker as the object, while leaving the grammatical function of the genitive case argument undecided. Lin's (2000) analysis as shown in Table 2 suggests that Pazeh language might have a nominative-accusative alignment system, but the framework she used appears to have disadvantages in presenting a comprehensive account of how the morphosyntactic operation works in voice alternations. Under the LFG framework, my analysis in the following sections shows that the distribution of three markers (i.e. *ki* re-analysed as absolutive case, *ni* as ergative case and *u* and *a* as oblique case) provides evidence that syntactically Pazeh has an ergative alignment system as shown in Table 3.

Table 2. Lin (2000: 123–124): Pazeh phrasal markers

Nominative	Genitive	Accusative	Locative
ki	ni	u/a	di

Table 3 Case markers and grammatical functions of Pazeh in this paper[3]

Form	ki	ni	u/a
Grammatical Functions	SUBJ	OBJ	OBL_{θ}
f-str info	(↑ CASE) = ABS	(↑ CASE) = ERG	(↑ CASE) = OBL

4 Ergative pattern: demotion and promotion of an argument

In the LFG framework (Bresnan & Kanerva, 1989; Bresnan & Mchombo, 1987; Nordlinger & Bresnan, 2011), it is uncontroversial that the argument preceded by the absolutive case marker *ki* is the syntactic subject (SUBJ). For instance, in a semantically and syntactically intransitive clause, as in (8) below, *ki* marks the sole core argument, and it also consistently marks the core argument selected by the designated voice as shown in (9) and (10). In other words, the voice morphology, as mentioned in the previous section, triggers the selection of a privileged argument to occupy the SUBJ slot, marked by the absolutive marker.

The grammatical functions of the non-subject arguments are not as obvious as the subject argument, yet the different distribution of phrasal markers *u* and *ni* reveals that the non-subject arguments in actor voice as in (8) and the one in undergoer voice as in (9) have different grammatical functions.

(8) liaka m-angit lia ki rakihan
then.ADV STAT-cry already.ADV ABS child.DEF
'Then the child cried.'

[3] To focus on the topic addressed in this paper, I will not discuss the grammatical function and the case marking of the previously labelled locative case.

(9)	mu-baket	(a)	wazu	ki	saw
	AV-hit	OBL	dog.INDEF	ABS	person.DEF
	'The person hit a dog.'				

(10)	baked-en	ni	Sabung	ki	rakihan
	hit-UV	ERG	person.name	ABS	child.DEF
	'Sabung hit the child.'				

(Li & Tsuchida, 2001: 42–44, 80–81; re-glossed by the author of this paper)

On the one hand, the marker *u* and *a* are used to denote the non-subject argument in actor voice clauses as in (9) and (11) for an indefinite patient. On the other hand, the marker *ni* denotes the non-subject argument in undergoer voice clauses as in (10) and (12) for a definite agent. The differences in the definiteness of the non-subject arguments in actor voice and undergoer voice suggest a difference in the semantic transitivity of the two voice types. Based on the transitivity parameter defined by Hopper and Thompson (1980), indefiniteness of the patient argument in actor voice clauses implies a less individuated or less affected argument, whereas the status of the agent arguments in undergoer voice clauses shows the opposite, always definite. The differences in the degrees of semantic transitivity support the hypothesis that an actor voice verb forms an extended intransitive construction where the non-subject argument is denoted by an oblique marker, and an undergoer voice verb forms a transitive construction, using a marker different from the one in actor voice.

(11)	me-ken	ki	balan	u	alaw
	AV-eat	ABS	cat.DEF	OBL	fish.INDEF
	'The cat ate fish.'				

(Lin, 2000: 102)

(12)	kan-en	ki	alaw	ni	balan
	eat-UV	ABS	fish.DEF	ERG	cat.DEF
	'The cat ate the fish.'				

(Lin, 2000: 132)

The distribution of phrasal markers for non-subject arguments in the instrumental voice clauses provides a further clue to clarify the status of the ergative case marked agent and oblique case marked patient. Similar to undergoer voice, the agent in the instrumental voice clause shown in (13) is also marked by an ergative marker *ni*. But different from undergoer voice, instrumental voice affixation not only has an effect on the selection of the subject argument, but also increases the valency of the verb from two to three. The instrument is marked by the absolutive marker, taking the SUBJ positive. Ergative case agent has a definite reading, but the patient argument marked by the oblique marker, is found to be indefinite. In topicalised clauses, the agent in instrumental voice and the agent in the undergoer voice as shown in (14) and (15) respectively, can both take the topic position, whereas empirical data show that there is no occurrence of topicalised indefinite patient in actor voice or instrumental voice. This also means that it is only possible to topicalise a patient argument when the clause is in undergoer voice. Topicalization[4] here works as a diagnostic measure to test the core status of non-subject arguments.

(13)	saa-baket	(a)	wazu	ni	rakihan	ki	patakan
	IV-hit	OBL	dog.INDEF	ERG	child.DEF	ABS	bamboo.DEF

‘The child used the bamboo stick hitting a dog.’

(Li &Tsuchida, 2001: 43)

(14)	yaku	ka,	sa-ken-an	ki	salaman	a	sumay
	1SG.FREE	TOP	IV-eat	ABS	bowl.DEF	OBL	rice.INDEF

‘I, used the bowl eating rice.’

(Lin, 2000: 135)

(15)	ita	ka,	ka-ken-en	di	laladan	ki	sumay
	1PL;INC.FREE	TOP	RED-eat-UV	P	table.DEF	ABS	rice.DEF

‘We all (inclusive), are going to eat the dishes at the table.’

(Lin, 2000: 128)

Semantic transitivity, phrasal markers and syntactic manipulation in topicalised clauses have all provided supporting information showing that the syntactic system of Pazeh is not as symmetrical as the system of Indonesian-type languages, such as Balinese (Arka, 2003). The GF for arguments marked by *u* or *a* is OBL_{θ}, and the most likely GF to be assigned to the argument marked by a core marker, *ni* will be OBJ. Under this analysis, the hallmarks of a Philippine-type voice system can be seen from the interaction of voice affixations and an ergative system with parallelism between *f-str* and *a-str* as shown in (16), (18) and (20) below.

With common typological notations, most transitive verb roots take two arguments, A for agent, P for patient as demonstrated in (16a). In line with the

[4] Topicalization has been used as a diagnostic test to examine the syntactic status of the arguments in other Austronesian language of Taiwan (Teng, 2008: 149–152).

notion of core/non-core[5] addressed by Arka (2003, 2019), I use angle brackets in the examples below to indicate the distinction between the core and the oblique arguments. As seen in (16) and the sentence (9), when actor voice prefixation takes place, the agent will be automatically selected to occupy the SUBJ position and linked to the pivot. The patient is demoted to oblique position and has obligatory indefiniteness reading. In (16b), an agent in the *sem-str* is mapped onto the a-subject in the *a-str*, which is realised as the GF-SUBJ. By contrast, as the voice alternation changes the valence of the predicate, a patient argument is turned into a non-core argument with empty termhood (Arka, 2003: 119–124), and it is realised as an OBL argument in the *f-str*. In the sense that the patient argument is demoted to the oblique position and read as indefinite, the actor voice construction as in (9) and (11) is considered to show an antipassive behaviour. The full lexical entry for the final verb form is presented in (16c) and the resulting f-structure of sentence (9) is shown in (17).

(16) a. -baket ‘hit < A P >’

pivot

b. mu-baket ‘AV.hit << S_A > < P > >’

GF:	SUBJ	$OBL_{PATIENT}$
a-str:	< a-subject >	< __ >
sem-str:	Agt (l-subject)	Pt (l-object)
f-str info:	↑ABS	↑OBL
marker:	*ki*	*u/a*

c. *mubaket* (↑PRED) = ‘hit << S_A > < P >>’
(↑VOICE-TYPE) = ACTOR
(↑SUBJ) = (↑PIVOT)
(↑SUBJ)σ= (↑A)
(↑SUBJ) =↓
(↓DEF) =c +
(↑OBL)σ= (↑P)
(↑OBL) =↓
(↓DEF) =c –
~(↑OBJ)

[5] The selecting properties for the core/non-core distinction include (but not limited to) case marking by the phrasal markers and topicalisation with voice constraint. For the discussion in this paper, I will only refer to these two properties.

(17)

$$\left[\begin{array}{ll} \text{PRED} & \text{‘hit <(}\uparrow\text{SUBJ) (}\uparrow\text{OBL)>’} \\ \text{OBL} & \left[\begin{array}{ll} \text{PRED} & \text{‘dog’} \\ \text{CASE} & \text{OBL} \\ \text{DEF} & - \end{array}\right] \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{‘person’} \\ \text{CASE} & \text{ABS} \\ \text{DEF} & + \end{array}\right] \\ \text{VOICE} & \text{ACTOR} \end{array}\right]$$

In undergoer voice, the patient is selected by the pivot, whereas the agent is kept at the argument structure as in (18), assigned with a distinct core marker for its status. In (18b), an agent is mapped onto the a-subject not a-object in the a-str. In the concept of parallel structures adopted by Arka (2003: 122), the l-subject is firstly mapped onto the a-subject, and in the next phase, the Agent-a-subject is encoded as the surface OBJ in the *f-str*. As for the argument selected by the pivot, a patient is the l-object mapped to a-object and realised as the surface SUBJ. In other words, undergoer voice clauses as in (10) and (12) remain transitive. The lexical entry is presented in (18c) and the f-structure of sentence (10) is shown in (19).

(18) a. -baket ‘hit < A P >’

pivot

b. baked-en ‘UV.hit < A P >’

GF:	OBJ	SUBJ
a-str:	< a-subject	a-object >
sem-str:	Agt (l-subject)	Pt (l-object)
f-str info:	↑ERG	↑ABS
marker:	*ni*	*ki*

c. *bakeden* (↑PRED) = ‘hit < A P >’
(↑VOICE-TYPE) = PATIENT
(↑SUBJ) = (↑PIVOT)
(↑SUBJ)σ= (↑P)
(↑SUBJ) =↓
(↓DEF) =c +
(↑OBJ)σ= (↑A)
(↑OBJ) =↓
(↓DEF) =c +

(19)

$$\left[\begin{array}{ll} \text{PRED} & \text{'hit <(}\uparrow\text{SUBJ) (}\uparrow\text{OBJ)>'} \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'Sabung'} \\ \text{CASE} & \text{ERG} \\ \text{DEF} & + \end{array}\right] \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'rakihan'} \\ \text{CASE} & \text{ABS} \\ \text{DEF} & + \end{array}\right] \\ \text{VOICE} & \text{PATIENT} \end{array}\right]$$

As for instrumental voice shown in (20) below, the voice affixation has an applicative effect on the re-structuring of *a-str* as in (20b). The voice morphology adds an instrument to the base and demotes the patient to OBL. This can be explained by the fact that the ergative pattern shows low tolerance of double non-agent arguments. As seen from (14), the agent is kept as a core argument, but the patient is demoted and marked by an oblique marker, indicating an indefinite reading as shown in (20c).

(20) a. -baket 'hit < A P >'

pivot

b. saa-baket 'IV.hit << A 'hit.with INST >' < P > >'

GF:	OBJ	SUBJ	$\text{OBL}_{\text{PATIENT}}$
a-str:	< a-subject	non-a-subject >	< __ >
sem-str:	Agt (l-subject)	Inst (l-oblique)	Pt (l-object)
f-str info:	↑ERG	↑ABS	↑OBL
marker:	*ni*	*ki*	*u/a*

c. *saabaket* (↑PRED) = 'hit < < A 'hit.with INST >' < P >>'
(↑VOICE-TYPE) = INSTRUMENTAL
(↑SUBJ) = (↑PIVOT)
(↑SUBJ)σ= (↑INST)
(↑SUBJ) =↓
(↓DEF) =c +
(↑OBJ)σ= (↑A)
(↑OBJ) =↓
(↓DEF) =c +
(↑OBL)σ= (↑P)
(↑OBL) =↓
(↓DEF) =c -

In other words, if the system itself were syntactically symmetrical, the demotion of P would not take place. The patient would have been kept as an OBJ. In short, the visualisation of the alignment system presented in (21) below indicates that syntactically the system is not symmetrical, and it shows an ergative pattern for clauses containing non-pronominals. The next question is whether Pazeh has a split alignment system for pronominals. In this regard, another piece of evidence is found for this ergative hypothesis from the pronominal paradigm in section 5.

(21)

<table>
<tr><td></td><td></td><td>Pivot</td><td></td><td></td></tr>
<tr><td>ITR.AV</td><td><</td><td>S_A</td><td>|</td><td>P ></td></tr>
<tr><td>TR.UV</td><td>< A</td><td>P</td><td>></td><td></td></tr>
<tr><td>TR.IV</td><td>< A</td><td>INST</td><td>|</td><td>P ></td></tr>
<tr><td>GF:</td><td>OBJ</td><td>SUBJ</td><td></td><td>$OBL_{PATIENT}$</td></tr>
<tr><td>f-str info:</td><td>↑ERG</td><td>↑ABS</td><td></td><td>↑OBL</td></tr>
<tr><td>marker:</td><td>ni</td><td>ki</td><td></td><td>u/a</td></tr>
</table>

5 Further evidence from pronominal paradigm

Pazeh has a distinct set of personal pronouns for the agent in non-actor voices as shown in Table 4 below. For instance, the first-person singular form for the non-subject core argument is *naki*, as in (22). This form is a specifically used to denote a non-pivot argument, and it is different from the form for the patient in undergoer voice, as in (23).

(22) kan-en naki dadua lia ki dadas
eat-UV 1SG.ERG all.ADV already.ADV ABS potato.DEF
'I ate all the potatoes.'

(Li &Tsuchida, 2001: 140)

(23) riud-en ni saw (y)aku
pinch-UV ERG person.DEF 1SG.ABS
'The person pinched me.'

(Li &Tsuchida, 2001: 254)

Table 4. Pazeh pronouns (Li &Tsuchida, 2001: 34)

	Agent in UV/IV	Agent in AV
1SG	*naki*	*(y)aku*
1PL;INCL	*nita*	*(i)ta*
1PL;EXCL	*niam*	*(ya)mi*
2SG	*nisiw*	*(i)siw*
2PL	*nimu*	*(i)mu*
3SG	*nimisiw*	*(i)misiw*
3PL	*namisiw*	*(ya)misiw*

For the agent in actor voice as in (24), the first-person singular form switches to *(y)aku* when it becomes the SUBJ or selected by the pivot. In comparison, the agent in the actor voice construction has the same form as the patient in the undergoer voice structure, whereas the SUBJ form is different from the second core argument in the undergoer voice.

(24)	m<in>eken	(a)	sumay	(y)aku
	AV-PERF-eat	OBL	rice.INDEF	1SG.ABS

‘I have eaten rice.’

(Li &Tsuchida, 2001: 140)

In sum, this pattern is in line with the ergative analysis, indicating that there is no split alignment system for the pronominals in Pazeh. The visualization of the ergative alignment for pronominals can be seen from (25) below. In addition to the pronominal forms shown in the examples, the non-pronominal arguments marked by the phrasal markers also demonstrate the corresponding case-marking and the grammatical functions as laid out in Table 3 in section 3. These examples show that Pazeh independent clauses are morphologically symmetrical for the voice affixation but syntactically ergative.

(25) Pivot

ITR.AV < S_A | P >

TR.UV < A P >

Pronominal form naki (y)aku

6 Concluding remarks

In this paper, I present evidence to show how the voice affixation is performed in Pazeh and how the morphological affixation interacts with a syntactically ergative system, triggering promotion and demotion of an argument. The evidence from semantic transitivity, topicalization and the distribution of phrasal markers in actor voice, undergoer voice and instrumental voice all indicate that Pazeh has an ergative alignment system and there is no split alignment system for the pronominals.

Under the LFG framework, the analysis in this paper reveals that there is a mismatch in the so-called symmetricality of Austronesian voice system in Pazeh. The mismatch lies between the symmetricality of voice affixation and the non-symmetricality of the syntactic system. Overlooking this mismatch would easily lead to confusion with the type of alignment system of Pazeh. A deeper investigation into the syntactic properties instead of the surface morphology breaks the myth that Pazeh voice system is symmetrical as a whole.

My findings also indicate that, the issue of Austronesian voice system can be well dealt with within the framework of LFG, by acknowledging the fact that symmetricality does exist in the voice system of Pazeh, but only at the morphological level, not at the syntactic level. Misleading terminology for the case marker and grammatical labels used by previous studies are avoided within LFG framework.

7 References

Arka, I. Wayan. (2003). *Balinese morphosyntax: A lexical-functional approach*. Canberra, Australia: Pacific Linguistics.

Arka, I. Wayan. (2019). Grammatical relations in Balinese. In Alena Witzlack-Makarevich & Balthasar Bickel (eds.), *Argument selectors: A new perspective on grammatical relations* (pp. 257–299). John Benjamins Publishing Co. doi:10.1075/tsl.123.08ark

Bresnan, Joan, & Kanerva, Jonni M. (1989). Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry*, *20*(1), 1–50. Retrieved from https://www.jstor.org/stable/4178612

Bresnan, Joan, & Mchombo, Sam A. (1987). Topic, pronoun, and agreement in Chicheŵa. *Language*, *63*, 741–782. doi:10.2307/415717

Falk, Yehuda N. (2006). *Subjects and universal grammar: An explanatory theory*. New York: Cambridge University Press.

Foley, William A., & Van Valin, Robert D. (1984). *Functional syntax and universal grammar*. New York: Cambridge University Press.

Himmelmann, Nikolaus P. (2002). Voice in western Austronesian: An update. In Fay Wouk and Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems* (pp. 7–16). Canberra,

Australia: Pacific Linguistics.
Himmelmann, Nikolaus P. (2005). The Austronesian language of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus. P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar* (pp. 110–181). New York, NY: Routledge.
Hopper, Paul J., & Thompson, Sandra (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251–299. doi:10.2307/413757
Li, Paul Jen-Kuei (李壬癸), & Tsuchida, Shigeru (eds.). (2001). *Pazih dictionary*. *Language and Linguistics, Monograph Series No. A2*. Taipei, Taiwan: Academia Sinica.
Li, Paul Jen-Kuei (李壬癸), & Tsuchida, Shigeru (eds.). (2002). *Pazih texts and songs*. Taipei, Taiwan: Academia Sinica.
Lin, Yin-Chin (林英津). (2000). *bā zé hǎi yǔ [The Pazeh language]*. Taipei, Taiwan: Yuanliu Publishing Company.
Manning, Christopher D. (1994). *Ergativity: Argument structure and grammatical relations*. (PhD dissertation, Stanford University, Stanford, UK). Retrieved from https://nlp.stanford.edu/~manning/dissertations/Manning-Christopher-ergativity-1994.pdf
Manning, Christopher D. (1996). *Ergativity: Argument structure and grammatical relations*. Stanford: Center for the Study of Language and Information.
Nordlinger, Rachel, & Bresnan, Joan (2011). Lexical-Functional Grammar: Interactions between morphology and syntax. In Robert D. Borsley & Kersti Börjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar* (pp. 112–140). doi:10.1002/9781444395037.ch3
Ross, Malcolm, & Teng, Stacy Fang-Ching. (2005). Formosan languages and linguistic typology. *Language and Linguistics*, *6*(4), 739–781. Retrieved from http://hdl.handle.net/1885/84418
Teng, Stacy Fang-Ching. (2008). *A reference grammar of Puyuma*. Canberra, Australia: Pacific Linguistics.
Zeitoun, Elizabeth, & Huang, Lillian M. (1997). Toward a typology of tense, aspect, and modality in the Formosan languages: A preliminary study. In Chiu-Yu Tseng (ed.), *Chinese Language and Linguistics IV: Typological Studies of Languages in China* (pp. 595–618). Taipei, Taiwan: Academia Sinica.