

Texas Linguistic Society IX:
The Morphosyntax of
Underrepresented Languages

Frederick Hoyt,
Nikki Seifert,
Alexandra Teodorescu,
and Jessica White
(vol. eds.)
and Stephen Wechsler (series ed.)

CENTER FOR THE STUDY
OF LANGUAGE
AND INFORMATION

Contents

- 1 **Case in Ergative Languages and NP
Split-Ergativity 1**
EDITH ALDRIDGE
- 2 **The Definite Article and Possessive Marking in
Amharic 21**
DOROTHEE BEERMAN AND BINYAM EPHREM
- 3 **How Universal is the Pragmatic Detachability
Scale: Evidence from Texas German Discourse
Markers 33**
HANS BOAS AND HUNTER WEILBACHER
- 4 **Rule Replacement in Lingít (Tlingit): The
Importance of Morphosyntax to Morphophonology 59**
SETH CABLE
- 5 **Number Marking in Maltese Nouns 79**
KARINE DAVID
- 6 **A Domain-Based Approach to 2P Clitics in
Pashto 89**
ASCANDER DOST
- 7 ***-ing* as an Agreement Marker in African American
English: Implications for Acquisition 111**
LISA GREEN

- 8 *Steal Me an Apple: Version in Georgian* 125
OLGA GUREVICH
- 9 **Denominal Verbs and Noun Incorporation:
Uto-Aztecan Evidence of a Unified Syntactic
Account** 145
JASON D. HAUGEN
- 10 **The Syntax of Agreement in Bantu Relatives** 167
BRENT HENDERSON
- 11 **Numeral Classifiers in Lhiimaqalhqama'** 185
SUSAN SMYTHE KUNG
- 12 **Argument Binding and Morphology in Chichewa** 203
SAM MCHOMBO
- 13 **Malagasy Instrumental Nominalizations** 223
DIMITRIS NTELITHEOS
- 14 **A Reanalysis of Nonemphatic Pronouns in
Dagbani** 239
TRISTAN PURVIS
- 15 **What Sorani Kurdish Absolute Prepositions Tell
Us About Cliticization** 265
POLLET SAMVELIAN
- 16 **Appalachian English *They*-Existentials** 287
ERIKA TROSETH
- 17 **Converb Constructions in Darma — A
Tibeto-Burman Language** 299
CHRISTINA M. WILLIS

Case in Ergative Languages and NP Split-Ergativity

EDITH ALDRIDGE

1 Introduction

A long-standing puzzle in the study of ergative languages is the phenomenon of NP split-ergativity. In the Pama-Nyungan language Dyirbal, case-marking on 3rd person arguments follows an ergative-absolutive pattern. Transitive subjects take the ergative suffix *-Ngu*, while transitive objects and intransitive subjects are morphologically unmarked.

- Dyirbal (Dixon 1994:161)
- (1) a. *yabu* *banaga-n^yu*
 mother.Abs return-Nonfut
 “Mother returned.”
- b. *Numa* *yabu-Ngu* *bura-n*
 father.Abs mother-Erg see-Nonfut
 “Mother saw father.”

1st and 2nd person pronouns, on the other hand, exhibit a nominative-accusative pattern. Objects take the accusative suffix *-na*, while subjects are morphologically unmarked.

- Dyirbal (Dixon 1994:161)
- (2) a. *Nana* banaga-n^yu
we.Nom return-Nonfut
“We returned.”
- b. *n^yurra* banaga-n^yu
you.Pl.Nom return-Nonfut
“You all returned.”
- c. *n^yurra* Nana-**na** bura-n
you.Pl.Nom we-Acc see-Nonfut
“You all saw us.”
- d. *Nana* *n^yurra-na* bura-n
we.Nom you.Pl-Acc see-Nonfut
“We saw you all.”

In other respects, however, Dyirbal is a syntactically ergative language. For example, clausal coordination takes place on an absolutive pivot, regardless of case-marking. In (3), the shared argument is the absolutive NP in both conjuncts: transitive object of the first conjunct and intransitive subject of the second conjunct.

- Dyirbal (Dixon 1994:162)
- (3) [**Numa** yabu-Ngu buran] [banagan^yu]
father.Abs mother-Erg saw returned
“Mother saw father and he returned.”

In (4), an example with pronominal arguments, the shared argument is likewise the transitive object of the first clause and the intransitive subject of the second clause.

- Dyirbal (Dixon 1994:162)
- (4) [*n^yurra* **Nana-na** buran] [banagan^yu]
you.Pl.Nom we-Acc saw returned
“You all saw us and we returned.”

Another Pama-Nyungan language Warlpiri exhibits a different type of NP split. The split in this case is between free and bound forms. NPs are marked with ergative and absolutive case. But agreement clitics, which double these arguments, show a nominative-accusative pattern. In the transitive clause in (5a), the 2nd person subject is marked with ergative case; in the intransitive clause in (5b), this subject is marked absolutive. However, the agreement marker takes the same form in both cases. (5a) and (5c)

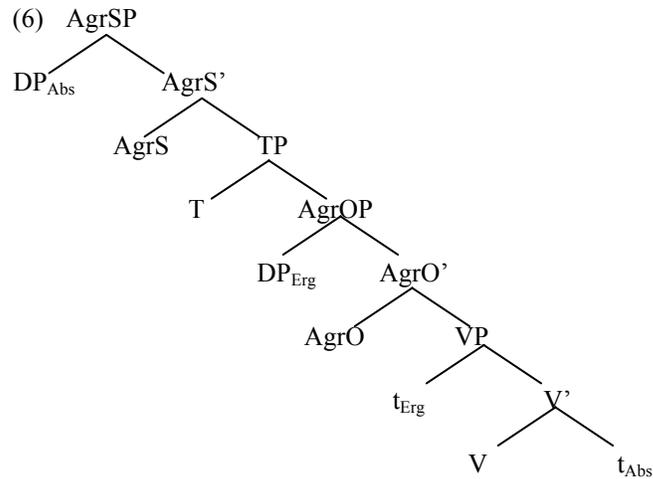
show that a transitive object receives the same absolutive case-marking as a intransitive subject, while the agreement clitics are different.

- Warlpiri (Bittner & Hale 1996:23)
- (5) a. *Nyuntulu-rlu* **ka-npa-ju** *ngaju* *nya-nyi*.
 you-Erg Pres-2s-1s me.Abs see-Nonpast
 “You see me.”
- b. *Nyuntu* **ka-npa** *parnka-mi*
 2s.Abs Pres-2s .Subj run-Nonpast
 “You are running.”
- c. *Ngaju* **ka-rna** *parnka-mi*.
 me.Abs Pres-1s run-Nonpast
 “I am running.”

This paper proposes an analysis of the two types of NP split-ergativity illustrated above. The main question to be addressed is whether the split is merely a morphological phenomenon or whether a separate syntactic mechanism is needed to account for it. I show that the split in case-marking is merely morphological and can be subsumed straightforwardly under a syntactic account of ergativity, assuming the late insertion model of Distributed Morphology. However, this is only possible under an analysis of ergativity which departs from the traditional approaches.

2 Challenge for Traditional Approaches to Ergativity

NP split-ergativity presents a serious challenge to traditional approaches to ergativity. Under such analyses, ergative and absolutive cases are taken to reflect grammatical functions and as such are uniformly associated with a particular functional category. In one approach, absolutives are treated as subjects and receive their case from the nominative-assigning functional projection (Murasugi 1992; Campana 1992; Bittner 1994; Bittner & Hale 1996; Manning 1996; Ura 2000). For example, Murasugi (1992) proposes that absolutive DPs move to [Spec, AgrSP] to check case, while ergatives check their case in [Spec, AgrOP].



This analysis faces difficulties accounting for paradigm in (2). For instance, the nominative pronouns in (2a) and (2b), as well as the accusative pronouns in (2c) and (2d), would move to [Spec, AgrSP], since they are syntactic absolutes. But the structure in (6) cannot determine when the morphological case on the pronoun should be null or should take the accusative suffix.

Levin & Massam (1985), Bobaljik (1993), and Laka (1993) take the opposite approach and propose analyses which treat the absolute as an object rather than the subject. In transitive clauses, ergative case is essentially equated with nominative and absolute with accusative. Therefore, ergative case is checked in [Spec, AgrSP], while absolute case is checked in the lower [Spec, AgrOP]. The problem presented by the paradigm in (2) is essentially the same as for the previous analysis. In this case, intransitive subjects and transitive objects will all move to [Spec, AgrOP]. But, again, no mechanism is in place for distinguishing when the case assigned by this functional head is absolute or accusative.

3 Subject in Ergative Languages

The main shortcoming of the traditional approaches to ergativity is the assumption that case-marking mirrors grammatical function. In this section, I show that neither absolutes nor ergatives behave uniformly as subjects. In most ergative languages, the ergative DP functions as the subject of a transitive clause, while absolutes exhibit subject-like behavior in intransitive contexts (Anderson 1976, Manning 1996, Ura 2000, Legate 2003, Aldridge 2004, and others).

Tagalog

- (17) Nag-ba-balak si Maria-ng
 Intr.Perf-Red-plan Abs Maria-Lk
 [PRO tulung-an si Pedro]
 (Erg) help-App Abs Pedro
 “Maria is planning to help Pedro.”

If the embedded clause is intransitive, however, then PRO appears in what would be the absolutive slot.

Tagalog

- (18) Gusto ni Maria-ng
 want Erg Maria-Lk
 [PRO b-um-ili ng libro]
 (Abs) -Intr.Perf-buy Obl book
 “Maria wants to buy a book.”

To summarize the discussion so far, it is the ergative DP which typically functions as the subject of a transitive clause, while the absolutive takes this role only in intransitive contexts. This raises the question, then, of what an absolutive is, if it is not a grammatical function, per se. Aldridge (2004 and 2005) has proposed that the most salient syntactic characteristic of absolutives in syntactically ergative languages is that they are the only DPs eligible to undergo A'-movement. (19) shows that absolutives, but not ergatives, can be relativized.

W. Greenlandic Eskimo (Manning 1996:84)

- (19) a. nanuq Piita-p tuqu-ta-a
 polar.bear.Abs Piita-Erg kill-Tr.Part-3sg
 “a polar bear killed by Piita”
 b. *angut aallaat tigu-sima-sa-a
 man.Abs gun.Abs take-Perf-Rel.Tr-3sg
 “the man who took the gun”

In order to relativize on an external argument, the clause must be anti-passivized and the external argument given absolutive status.

Dyirbal (Dixon 1994:169-170)

- (20) yabu_i [e_i bural-nga-ngu nguma-gu]
 mother.Abs see-AP-Rel.Abs father-Dat
 banaga-nyu
 return-Nonfut
 “Mother, who saw father, was returning.”

(21) shows the same facts for *wh*-movement in Jacaltec Mayan. The object absolutive can be extracted directly from the transitive clause in (21a). But the verb must be antipassivized to extract the subject, as shown in (21b).

Jacaltec Mayan (Craig 1977:14)

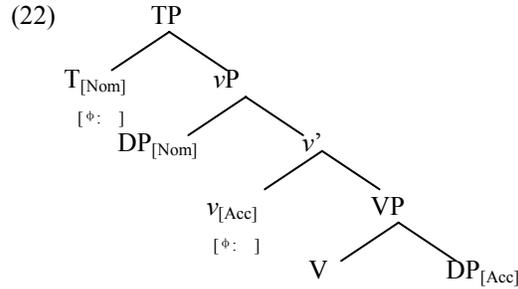
- (21) a. mac xawila
 whom you.saw
 “Whom did you see?”
 b. mac xcach mak-**ni**
 who you hit-AP
 “Who hit you?”

In Sections 4 and 5, I present the analysis of case and agreement in Warlpiri and Dyirbal. In Section 6, I return to the issue of syntactic ergativity and propose an analysis of the absolutive restriction on A'-extraction.

4 *v*-Type Ergativity¹ and the Bound/Free Split

The primary conclusion of the preceding section is that absolutives function as subjects only in intransitive clauses and should be treated as objects in transitive contexts. This section presents a formal analysis of this split in subject and object behavior. The proposal is grounded in recent Minimalist theory, as proposed by Chomsky (2000, 2001a, 2001b). In a transitive clause in an accusative language, T values nominative case on the subject DP, while transitive *v* values accusative case on the direct object. In languages with subject and object agreement, the ϕ -features of the object are copied to *v* and those of the subject are copied to T.

¹ Aldridge (2004) proposes that there are two types of syntactically ergative language: *v*-type, in which absolutive case-checking is shared by T and *v*; and T-type, in which absolutive case is checked uniformly by T. In the current paper, I discuss only *v*-type ergativity.

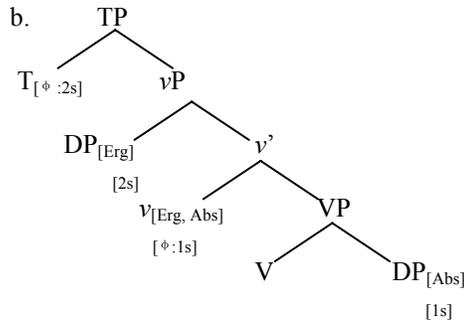


This system can be translated almost directly to ergative languages. The split in object and subject properties exhibited by absolutes is captured by merging the absolutive case feature in T in intransitive clauses and *v* in transitive clauses. Ergative case is inherent, assigned by transitive *v*.² Agreement takes place just as in (23): T copies the ϕ -features of the subject (the ergative DP in transitive clauses and the absolutive in intransitive clauses). *v* copies the ϕ -features of the absolutive object in transitive clauses.

To see how this case and agreement system works for Warlpiri, consider the transitive clause in (24). Transitive *v* carries an absolutive case feature, which it values on the object DP. It also copies the ϕ -features of the object and assigns inherent ergative case to the external argument. T does lack a case feature but copies the subject's agreement features.

Warlpiri (Bittner & Hale 1996:23)

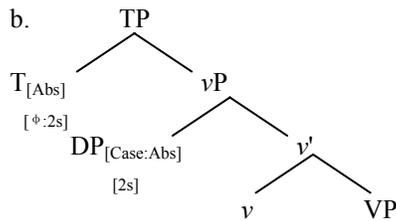
- (23) a. Nyuntulu-*rlu* ka-**npa-ju** *ngaju* nya-nyi.
 you-Erg Pres-2s-1s me.Abs see-Nonpast
 "You see me."



² Bittner and Hale (1996), Ura (2000), Legate (2003), Aldridge (2004), and others have similarly proposed that ergative case is inherent and not structural.

In an intransitive clause, ν has neither case nor agreement features. T assigns absolutive case to the subject and copies its ϕ -features.

- (24) a. Warlpiri (Bittner & Hale 1996)
Nyuntu **ka-npa** *parnka-mi.*
 2s.Abs Pres-2s .Subj run-Nonpast
 “You are running.”



NP split-ergativity between agreement and case-marking will be realized automatically, since the agreement features copied to T are those of subjects and the agreement features copied to ν are those of objects. We need only to designate that ϕ -features on T are replaced with forms from the nominative paradigm and those on ν are taken from the accusative paradigm.³

The question might be raised at this point as to how the syntax knows when T has a case feature and when it does not. The answer to this question is that this feature is freely assigned to T. It is an uninterpretable feature and as such must be checked prior to Spell-Out. It can only be checked when it enters into an Agree relation with an unvalued case feature on a DP in its c-command domain, as per Minimalist assumptions. This entails that the derivation only converges when T carries an absolutive case feature in intransitive clauses. This is because in transitive clauses, the object and subject are assigned case by ν and therefore do not have an unvalued case feature to check with T. Therefore, an uninterpretable absolutive case feature on T in a transitive clause is not checked, and the derivation crashes.

In intransitive clauses, however, T must have a case feature; otherwise, the unvalued case feature of the absolutive, which is not only unvalued but also uninterpretable, will go unvalued and unchecked and the derivation

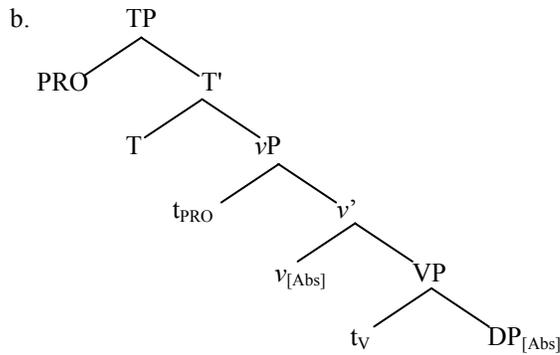
³ I am assuming the late insertion model of Distributed Morphology, as proposed by Marantz (1991), Halle and Marantz (1993), among many others. More will be said about this model in Section 5.

will crash. Therefore, the derivation converges exactly when T has an absolutive case feature in intransitive clauses.

The analysis of *v*-type ergativity presented above accounts for the subject properties observed in Section 3 in the following ways. In transitive clauses, it is the ergative DP which binds reflexives, because it resides in the highest A-position in the clause and c-commands all other arguments. In an antipassive, the external argument is assigned absolutive case by T, but it still resides in its base position and therefore can bind a reflexive inside VP. Ergative DPs function as imperative and hortative addressees, by virtue of the fact that they are external arguments.

The most interesting fact concerns nonfinite embedded clauses, particularly the availability of absolutive case for an object. This is accounted for in the analysis above, since it is *v* which assigns absolutive case in transitive clauses and not T.

- (25) a. Tagalog
 Bina-balak ni Maria-ng
 Tr.Prog-plan Erg Maria-Lk
 [PRO tulung-an si Pedro]
 (Erg) help-App Abs Pedro
 “Maria is planning to help Pedro.”



5 Person-based Split in Case-marking

The type of split-ergativity in Dyirbal must be treated differently from Warlpiri. Here, the split is manifested entirely in the case-marking system, with 1st and 2nd person pronominal arguments assigned nominative and accusative case, and 3rd person arguments marked ergative and absolutive.

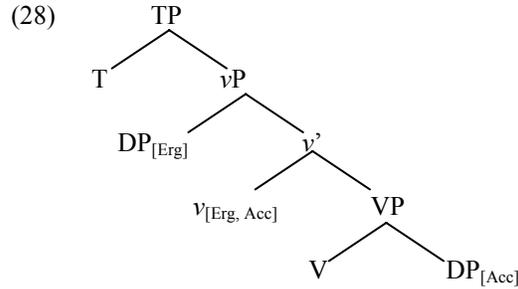
- Dyirbal (Dixon 1994:161)
- (26) a. *Nana* *n^yurra-na* *bura-n*
 we.Nom you.Pl-Acc see-Nonfut
 “We saw you all.”
- b. *Numa* *yabu-Ngu* *bura-n*
 father.Abs mother-Erg see-Nonfut
 “Mother saw father.”

The problem for the analysis of case-marking proposed in Section 4 is that absolutive case is assigned to both the absolutive and accusative objects in (26). Likewise, ergative case is assigned to both the ergative and nominative subjects in transitive clauses. Therefore, the proposal in Section 4 must be revised. First, note that there are three morphological cases for subjects and objects in Dyirbal (absolutive and nominative take the same form).

- (27) Summary of Dyirbal Person Split
- | | |
|--|-------------------|
| 3 rd person: | Erg = <i>-Ngu</i> |
| | Abs = NULL |
| 1 st /2 nd person: | Nom = NULL |
| | Acc = <i>-na</i> |

The first step, then, is to add a case feature to the inventory of cases assigned by T and ν . To this end, I adopt part of Legate’s (2003) analysis of case assignment in Warlpiri.⁴ The key aspect of this proposal is that the structural cases assigned by T and transitive ν are distinct. The case feature assigned by T remains absolutive, but the one carried by transitive ν is accusative. Inherent ergative case is still assigned by transitive ν to its specifier.

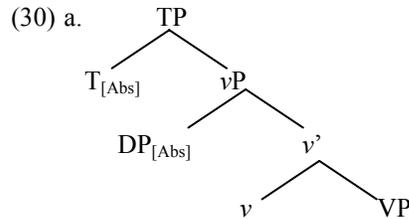
⁴ I do not, however, adopt every aspect of Legate’s proposal. One aspect I do not adopt is her idea that absolutive (nominative) case is always available on T, even in transitive clauses, when it is not assigned to a DP. This has the unattractive result of allowing an uninterpretable feature to survive until Spell-Out.



Prior to Spell-Out, the subject and objects in (26) will have the following case features.

- (29) a. Nana_[Erg] n^yurra_[Acc] bura-n
 we you.Pl see-Nonfut
 “We saw you all.”
- b. Numa_[Acc] yabu_[Erg] bura-n
 father mother-Erg see-Nonfut
 “Mother saw father.”

In an intransitive clause, *v* has no case feature, so T must assign absolutive case to the single argument.



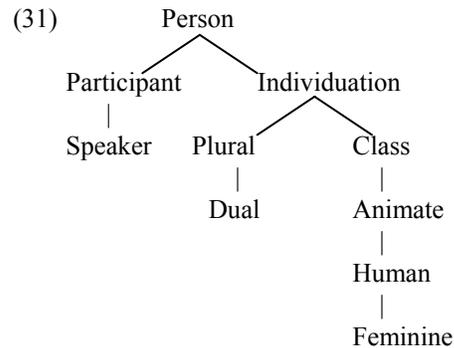
- b. yabu_[Abs] banaga-n^yu
 mother return-Nonfut
 “Mother returned.”
- c. Nana_[Abs] banaga-n^yu
 we return-Nonfut
 “We returned.”

The case features cannot be spelled out directly. This would result in all transitive objects being marked accusative and all transitive subjects taking the ergative suffix. This potential problem can be circumvented, however, by assuming late insertion within the framework of Distributed

Morphology. Recall first that 1st and 2nd person form a natural class, in that these are marked according to an accusative pattern, which is distinct from 3rd person, which follows an ergative pattern. This grouping is naturally captured in recent approaches to the organization of morphosyntactic features in Distributed Morphology, which views these features as being arranged according to a geometry.

Feature Geometry

(Adapted from Harley (1994), Harley & Ritter (2002))



This geometry groups 1st and 2nd person together under ‘participant’. ‘Speaker’ represents 1st person; 3rd person is represented by the absence of the ‘participant’ node and is therefore treated as a default. The morphological forms are inserted according to the following vocabulary items.

(32) Vocabulary Items

[Case:Acc, Part] ⇔ *-na*

[Case:Erg, Part] ⇔ NULL

[Case:Erg] ⇔ *-Ngu*

[Case] ⇔ NULL

The vocabulary items are referenced in the order given. This ensures that the more marked forms are spelled out first, followed by the default forms. Therefore, 1st and 2nd person objects will be given the accusative suffix *-na*. 1st and 2nd person subjects will be treated as nominative, which is null. All other transitive subjects will take the ergative suffix *-Ngu*. The remaining subjects and objects will be unmarked, as the default absolutive or nominative case.

The feature geometry in effect subsumes animacy hierarchies like those proposed by Silverstein (1976) and Dixon (1994). Silverstein’s insight is

while treating 3rd person as the default. Animate and Human are also grouped as natural classes and are treated as more marked than inanimate.

6 Syntactic Ergativity

In this section, I briefly propose an account of the main characteristic of syntactic ergativity: the absolutive restriction on A'-extraction. As discussed in Section 3, only absolutive DPs are eligible to undergo A'-movement in syntactically ergative languages.

- (35) a. W. Greenlandic Eskimo (Manning 1996:84)
- | | | |
|--------------------------------|-----------|------------------|
| nanuq | Piita-p | tuqu-ta-a |
| polar.bear.Abs | Piita-Erg | kill-Tr.Part-3sg |
| “a polar bear killed by Piita” | | |
- b. *angut aallaat tigu-sima-sa-a
- | | | |
|----------------------------|---------|----------------------|
| man.Abs | gun.Abs | take-Perf-Rel.Tr-3sg |
| “the man who took the gun” | | |

A straightforward analysis of this restriction is readily available, given the status of vP as a phase and the Phase Impenetrability Condition (Chomsky 2001b:5).

(36) Phase Impenetrability Condition (PIC)

The domain of a phase head is not accessible to operations, but only the edge is.

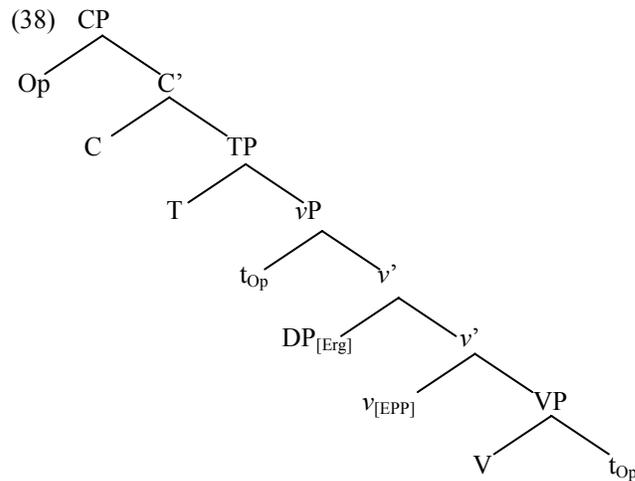
Specifically, only the head and specifier positions of vP are potential launching sites for movement to [Spec, CP] or beyond the clause. This means that VP-internal material must also pass through one of these positions, before leaving vP . Given that movement within Minimalist theory is always feature-driven, v must have an appropriate feature, typically an EPP feature, to draw a VP-internal constituent to its outer specifier. From this position in the edge of vP , the constituent in question is now accessible to a feature which will attract it to the next phase edge, i.e. [Spec, CP]. Direct movement from within VP to [Spec, CP] would violate the PIC.

It is assumed for English that EPP features are generated on v any time they are needed to facilitate movement. What I propose (see Aldridge 2004, 2005 for additional discussion) for syntactically ergative languages is that the appearance of EPP features on v is restricted.

(37) Transitivity and EPP

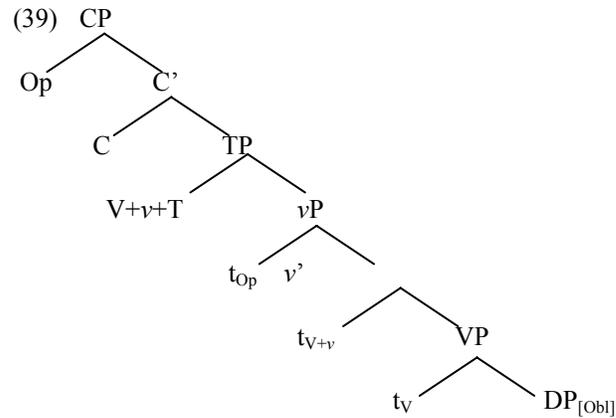
- ⇒ Transitive v has an EPP feature, drawing the absolutive DP to its outer specifier.
- ⇒ Intransitive, including antipassive, v has no EPP feature.

Extraction from a transitive clause is accounted for as follows. The EPP feature on transitive v attracts the VP-internal absolutive to the vP phase edge, making it the closest DP to C and ensuring that it is the DP which will be able to undergo A'-movement to [Spec, CP].



In an antipassive – which is intransitive – v does not have an EPP feature, so the DP merged in [Spec, vP] is now the closest DP to C. A VP-internal DP cannot be attracted without violating the Phase Impenetrability Condition. But the external argument, which is base merged in a specifier of v is eligible, accounting for the fact that agent extraction is possible only in antipassives.⁵

⁵ Extraction of an unaccusative subject is also direct, from inside VP to [Spec, CP]. Chomsky (2000, 2001a, 2001b) that unaccusative vP is a weak phase. Direct extraction from the domain of a weak phase does not violate the PIC.



The absolutive pivot in clausal coordination is accounted for similarly, assuming that zero-pronominalization in the second conjunct is fed by topicalization.

7 Conclusion

In this paper, I have presented an analysis of morphological and syntactic properties of *v*-type ergative languages. The most salient characteristic of this type of ergativity is that it is *v* which assigns case to absolutive objects in transitive clauses, while T assigns case to absolutives in intransitive contexts. This division of labor between the two functional heads in assigning absolutive case mirrors the non-uniform behavior of absolutive DPs with respect to subject and object properties.

Distinguishing transitive objects from both transitive and intransitive subjects is also what underlies the analysis of NP split-ergativity. Nominative/accusative agreement in Warlpiri is realized by copying subject ϕ -features to T and object ϕ -features to *v*, which essentially amounts to a direct mapping from argument structure to agreement and therefore need not be considered an aberrant pattern, even in an otherwise ergative language. The distinction between transitive and intransitive absolutives also plays a crucial role in accounting for the person-based split in Dyirbal, specifically by allowing for the assignment of three morphologically distinct cases. The feature geometry assumed in Distributed Morphology then isolates 1st and 2nd person and allows them to be treated distinctly from 3rd person in the morphological component.

The special role of transitive *v* is also seen in the account of the absolutive restriction on A'-extraction. EPP features can be merged only on

transitive (and not antipassive) *v*, with the result that only transitive objects and intransitive subjects are allowed to undergo A'-movement.

8 References

- Aldridge, E. 2004. *Ergativity and Word Order in Austronesian Languages*. Ph.D. dissertation, Cornell University.
- Aldridge, E. 2005. Phase Theory Account of Absolutive Extraction in Tagalog. *MIT Working Papers in Linguistics 49: Perspectives on Phases*, ed. M. McGinnis and N. Richards, 1-28.
- Anderson, S. 1976. On the Notion of Subject in Ergative Languages. *Subject and Topic*, ed. C. Li, 3-23. New York: Academic Press.
- Bittner, M. 1994. *Case, Scope, and Binding*. Dordrecht: Kluwer Academic Publishers.
- Bittner, M. and K. Hale. 1996. The Structural Determination of Case and Agreement. *Linguistic Inquiry* 27:1-68.
- Bobaljik, J. 1993. On Ergativity and Ergative Unergatives. *MIT Working Papers in Linguistics 19: Papers on Case and Agreement II*, 45-88.
- Campana, M. 1992. *A Movement Theory of Ergativity*. Ph.D. dissertation, McGill University.
- Chomsky, N. 2000. Minimalist Inquiries. *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, ed. Martin et al., 89-155. Cambridge, MA: MIT Press.
- Chomsky, N. 2001a. Derivation by Phase. *Ken Hale: A Life in Language*, ed. M. Kenstowicz, 1-52. Cambridge, MA: MIT Press.
- Chomsky, N. 2001b. Beyond Explanatory Adequacy. *MIT Occasional Papers in Linguistics 20*.
- Craig, C. 1977. *The Structure of Jacalteco*. Austin: University of Texas Press.
- Dixon, R.M.W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. 1994. *Ergativity*. Cambridge University Press.
- Halle, M. and A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. *The View from Building 20*, ed. K. Hale & J. Keyser, 111-176. Cambridge, MA: MIT Press.
- Harley, H. 1994. Hug a Tree: Deriving the Morphosyntactic Feature Hierarchy. *MIT Working Papers in Linguistics 21: Papers on Phonology and Morphology*, ed. A. Carnie & H. Harley, 289-320.
- Harley, H. and E. Ritter. 2002. Person and Number in Pronouns: A Feature-Geometric Analysis. *Language* 78.3:482-526.
- Laka, I. 1993. Unergatives that Assign Ergative, Unaccusatives that Assign Accusative. *MIT Working Papers in Linguistics 18, Papers on Case and Agreement I*, 149-172.

- Larsen, T. W. and W. M. Norman. 1979. Correlates of Ergativity in Mayan Grammar. *Ergativity: Towards a Theory of Grammatical Relations*, ed. F. Plank, 347-370. New York: Academic Press.
- Legate, J. 2003. *Warlpiri: Theoretical Implications*. Ph.D. Dissertation, MIT.
- Levin, J. and D. Massam. 1985. Surface Ergativity: Case/Theta Relations Reexamined. *Proceedings of NELS 15*, ed. S. Berman, J. Choe, J. McDonough, 286-301.
- Manning, Christopher. 1996. *Ergativity: Argument Structure and Grammatical Relations*. Stanford: Center for the Study of Language and Information.
- Marantz, A. 1991. Case and Licensing. In *Proceedings of ESCOL*, 234-253.
- Morphy, F. 1983. Djapu, A Yolngu Dialect. *Handbook of Australian Languages*, vol. 3, ed. R.M.W. Dixon and B.J. Blake, 1-188. Canberra: ANU Press.
- Murasugi, K. G. 1992. *Crossing and Nested Paths: NP Movement in Accusative and Ergative Languages*. Ph.D. dissertation, MIT.
- Payne, T. E. 1982. Role and Reference Related Subject Properties and Ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6, 1:75-106.
- Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. *Grammatical Categories in Australian Languages*, ed. R.M.W. Dixon, 112-171. Canberra: Australian Institute of Aboriginal Studies.
- Ura, Hiroyuki. 2000. *Checking Theory and Grammatical Functions in Universal Grammar*. New York: Oxford University Press.

The Definite Article and Possessive Marking in Amharic

DOROTHEE BEERMANN AND BINYAM EPHREM

1 Introduction

In this paper we are concerned with the grammar of noun phrases in Amharic, a Semitic language spoken in Ethiopia. Our theoretical frame is Head-Driven Phrase Structure Grammar (HPSG) and our focus is the Amharic definite article, a suffix which is homonymous with the 3rd singular possessive marker. In this paper we would like to account for the left edge effect of the Amharic definite marking as well as the distribution of definiteness marking in modified possessive noun phrases. Our suggestion is to promote definiteness from the head of the syntactic left-hand daughter, while, for example, person and number specifications, and also a POSS(essive) feature, are promoted from the head daughter; the latter is in accordance with the more general constraints embodied in the Head Feature Principle of HPSG.

Nouns in Amharic may carry affixes to indicate their gender, number, definiteness, and case. The language is at least partially agglutinative. Morphemes are suffixed to the noun in the order listed above (see also: Leslau 1995). Nominal modifiers and specifiers may host some of the nominal morphology, leading to patterns of nominal inflection throughout the noun phrase to be described below.

The paper is organized as follows: the second section describes possessive and definite marking while the third section gives a short overview of previous accounts of Amharic nominals and outlines some basic assumptions about noun phrases in HPSG. The fourth section suggests a constraint-based analysis of definite and possessive forms, and section five contains the conclusion.

2 Possessive and definite markers

2.1 The Possessive

Consider the following examples:

- (1) a. *məs'haf*
 'a book'
 b. *məs'haf-e*
 book -1per.sg. (poss)
 'my book'
 c. *yə-ənnə məs'haf*
 Gen- I book
 'my book'

Examples (1b) and (1c) show the two ways of constructing possessives in Amharic. In (1b) a possessive suffix is attached to the noun stem, that is, the morpheme */-e/* adds 1per.sg information, as it introduces a possessor, and *məs'haf* 'book' becomes 'my book'. The full array of possessive suffixes is shown in Table 1:

Singular		
	1	-e
	2.masc	-h
	2. fem	-š
	3.masc	-u/-w
	3. fem	-wa
Plural		
	1	-aččən
		-aččəhu
	3	-aččəw

Table 1. *Possessive suffixes in Amharic*

The second way of expressing possession is through the use of the genitive prefix /yə-/ (1c). It is attached to an independent pronoun or a noun and expresses ownership, but also part-whole relations and other asymmetric semantic dependencies between nominal entities. This latter construction has been treated, e.g., by den Dikken 2004, and will not further concern us here.

2.2 Definite Marking

Consider the following examples:

- | | |
|---|--|
| <p>(2) a. (and) zaf
 (one) tree
 ‘a tree’
 b. zaf-u
 tree- DEF/3masc.sg
 ‘the/his tree’</p> | <p>(3) a. wəša
 ‘a dog’
 b. wəša-w
 dog- DEF/3masc.sg
 ‘the/his dog’</p> |
|---|--|

Unlike in English, Amharic singular indefinite noun phrases do not need to be specified by a determiner, that is, bare singular nouns can serve as arguments and receive an indefinite interpretation. Indefinite noun phrases may also be construed with the help of the specifier *and* which literally means ‘one’, but when co-occurring with a noun often leads to an indefinite interpretation, as indicated in (2a). Definite singular nouns are derived via the suffixation of *u* or *w* to the singular noun, an alternation which is due to allomorphy: /-u/ goes to [-w] after a vowel, while /-u/ becomes [-u] after a consonant. As can be seen from Table 1 above, the definite article suffix and the possessive marker for third person masculine singular are homonyms. Depending on the context, however, it is not always the case that both meanings of a noun with a *u/w* suffix can be retrieved. If for example an object generally is not owned, such as a *river*, the possessive meaning of the noun becomes implausible, giving way to the definite interpretation only:

- | | |
|--|--|
| <p>(4) a. (and) wənz
 a river
 ‘a river’</p> | <p>b. wənz-u
 river-DEF
 ‘the river’</p> |
|--|--|

Another example of the unambiguous use of the *u*-suffix derives from the possessive marking of inherently feminine nouns like *lam* ‘cow’. As mentioned above, nouns in Amharic can be inflected for gender. The marker for

feminine is /-it/ (see also Leslau 1995) while the marker for a definite feminine noun is, just like its masculine counterpart, homophonous with the 3p.fem possessive marker, so that the following patterns arise:

- | | |
|--|--|
| (5) a. lam-wa
cow- DEF/3fem.sg
'the/her cow' | b. lam-it-uwa
cow-fem-DEF/3fem.sg.
'the/her cow' |
| c. lam-it-u
cow-fem-DEF/3masc.sg
'the/his cow' | d. lam-u
cow-3masc.sg
'his cow' |

In the case that only the definite suffix is attached to *lam*, deriving *lamu*, (5d), its meaning is only possessive, since in order to be definite an inherently feminine noun has to be specified for gender in one of the ways indicated in (5).

2.3 Definite Marking in an Extended Noun Phrase

In Amharic, NPs adjectives and quantifiers precede the noun. Let us consider the following examples:

- (6) a. tələk' wənz/bet
big river/house
'a big river/house'
- b. tələk'-u wənz/bet
big-DEF river/house
'the big river/house'
- c. tələk'-u (wənz- u)/(bet-u)
big-DEF (river-poss.3msg)/(house- poss.3msg)
'his big river/house'
- d. *tələk' (wənz-u)/(bet-u)
big river-poss.3msg/house-poss.3msg

In examples (6a)-(6d) we make use of the fact that the possessive marking of a landmark like a river is in most cases contextually odd, to highlight the fact that what might look like definite agreement on adjective and noun is in fact the obligatory definiteness marking of the adjective in a possessive noun phrase. The *-u* suffixed to the adjective is the definite article suffix while the *-u* suffix attached to the head noun is the possessive marker. Let us examine this point further. Consider (7):

- (7) a. tələk'-u bet-e
 big- DEF house-1sg.(poss.)
 'my big house'
 b. *tələk' bet-e
 big bet-1sg(poss.)

When *bet-e* 'my house' is modified by an adjective, the adjective obligatorily attaches the definite article suffix. A modified possessive noun phrase thus needs, also morphosyntactically, to be marked as definite.

Let us now consider the following NPs:

- (8) a. bet'am tələk'-u wənz
 very big-DEF river
 'the very big river'
 b. *bet'am-u tələk' wənz
 very-DEF big river
 c. *bet'am-u tələk'-u wənz
 very-DEF big- DEF river
- (9) a. hulət-u tələlək' ləḡ-očč
 two-DEF big.PL child-PL
 'the two big children'
 b. hulət-u tələlək' ləḡ-očč
 two-DEF big.PL child-PL
 'the two big children'
 c. hulət-u tələlək' ləḡ-očč-u
 two-DEF big.PL child-PL-3masc(poss.)
 'the two of his big children'
 d. *hulət tələlək'-u ləḡ-očč
 two big.PL-DEF child-PL

In (8a) the definite article is attached to the adjective, while it is ungrammatical to attach it to an intensifier (8b) or to both the adjective and the intensifier (8c). In the case of a quantifier (9), it is only the quantifier that takes the definite marker (9b). Finally, attaching the marker to other constituents will again lead to ungrammaticality (9d).

In summary, what we observe is that the Amharic definite article is suffixed to the head of the left-most constituent within the noun phrase. This explains that the intensifier, which is a specifier of the adjective modifying the noun, is not a possible host for the definite article, while an adjective or a quantifier, both heading their own projection, are.

Let us finally point out that Amharic demonstrative pronouns generally cannot co-occur with the suffixed definite article as, e.g., in (10a) or (10b) below:

- (10) a. *yəh betam tələk' * (tələk'-u) bet(-u)*
 Dem.sg very big *(big-DEF) house(-3msg(poss.))
 'this very big house (of him)'
- b. *ənziya hulət *(-u) tələlək' ləḡḡ-očč (-u)*
 Dem.pl two *(-DEF) big.PL child-PL (-3msg (poss.))
 'those two big children (of him)'

This is allegedly a reflex of the fact the *u/w* suffix is not an agreement marker but rather the morphological realization of the definite article. Assuming that multiple specifiers are disallowed in Amharic, the facts in (10) would follow.

Let us summarize:

- the morpheme /-u/ is a definite article which can be suffixed to nouns, adjectives, or quantifiers
- 3person possessive markers are homophonous with the definite article suffix
- left-hand daughters of possessive nouns, independent of their functional status, need to be morphologically marked as definite.
- Amharic NPs are marked for definiteness on the head of the left-most constituent.

3 The definite article

3.1 The Amharic Definite Article

The definite article in Amharic is a morphologically bound element, a fact that has triggered some theoretical disagreement concerning its functional status. Against Lyons (1999), who treats the Amharic definite article as a clitic, but with Wintner (2000), using HPSG, as well as den Dikken, using

the Minimalist Program (2004), we would like to treat the Amharic article as an inflectional suffix. We furthermore would like to assume that the noun is the semantic and syntactic head of the nominal phrase and that the indefinite determiner *and*, demonstratives, and quantifiers are selected by the nominal head. The definite article, however, is an inflectional suffix, added to the nominal stem by a lexical rule. In the following we describe how such a rule could look and its interaction with possessive marking on the one hand, and the left-edge effect of the Amharic definiteness marking on the other.

3.2 The Treatment of the Definite Article in HPSG

HPSG has so far been primarily concerned with English. In English the article is an independent word preceding the noun, which, in an HPSG grammar, is combined with the head through the SPECIFIER-HEAD SCHEMA (Pollard and Sag 1994). Without being able to explain its theoretical background, the essence of the SPECIFIER-HEAD SCHEMA may be characterized as follows: the head noun subcategorizes for its determiner, while the determiner may specify in its SPEC feature certain properties of the noun that it restricts. The SPECIFIER-HEAD SCHEMA thus comes very close to a formalization of the double-headedness of noun phrases. In the interaction of the constraints imposed on the values of the SPEC and SPR(specifier) features, nominal inflection with respect to, e.g., number, gender and definiteness can be expressed. With the extension of HPSG to languages where the expression of definiteness is part of the morphological component of the language, as for example in the Scandinavian languages (Hellan and Beermann 2005) and within the Semitic language family, for example for Modern Hebrew, linguists working within HPSG have suggested an account of definiteness as part of the lexical rule inventory of the language. Wintner (2000), for example, suggests to treat the definite marker in Modern Hebrew as an affix and to combine it with nouns, adjectives, and quantifiers in the lexicon. Here we will follow Wintner's approach and treat the Amharic definite suffix as an inflectional affix. However, different from Hebrew, Amharic marks definiteness only once and only on the head of the left daughter of a noun phrase. Departing from Wintner's word-to-word rule of definiteness marking in Hebrew, we will in the following develop a set of morphological rules that will try to accommodate the Amharic left-edge effect of definiteness marking. We will pay special attention to possessive noun phrases and the effects that arise from the homophony between the 3sg possessive markers and the obligatory definiteness marking of possessive phrases.

4 The Analysis

Wintner (2000) suggests the following Definite Lexical Rule to derive definite nominal words:

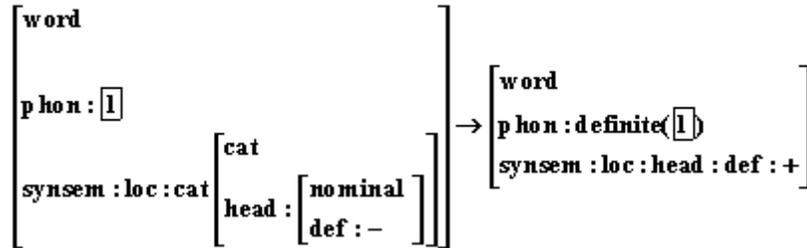


Figure 1. *Wintner's Definite Lexical Rule for Modern Hebrew* (Wintner 2000: 349)

Wintner's rule applies to indefinite nominal words which are nouns, adjectives, cardinals, and numerals, and returns definite nominal words. Note that Wintner uses a denotation different from ours: we separate features in a feature path by a period rather than by a double colon, the PHON feature is a list value, words are strings, and the output of an inflectional rule will be given by indicating morph-boundaries by hyphens in the PHON value, as shown in Figure 2. Also note that the Hebrew definite marker is a prefix while it is a suffix in Amharic. In our adaptation of Wintner's rule, the most substantial modification of Figure 1, however, lies in the fact that in Amharic, definiteness is not expressed as a head feature, instead the definiteness feature DEF is a sister to HEAD, for reasons to be explained shortly:

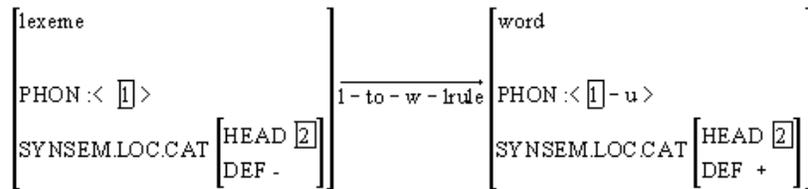


Figure 2. *Definite-lexeme-to-word-rule for Amharic*

Definiteness in Amharic is signalled on the head of the left-hand daughter independently of its categorial or functional status. To account for this

fact we assume that definiteness of a non-unary NP is propagated from the left-hand daughter to the mother node. HEAD-SPECIFIER and HEAD-MODIFIER RULE need to be rendered accordingly. In the following we only consider the HEAD-MODIFIER RULE in more detail.

In Figure 3 we use the HEAD-MOD SCHEMA to combine an adjective with the noun it modifies. In HPSG, it is the adjunct that selects the element it modifies. In Amharic, independent of whether we deal with possessive nouns or not, it is an adjective specified as [DEF +] that selects in its MOD feature a noun that is specified as [DEF -]. This means that definiteness will always reside in the left hand daughter (nonhead daughter). We suggest that the DEF value of the left-hand daughter is re-entered with the DEF value of the phrase as a whole. The latter fact is shown in Figure 3:

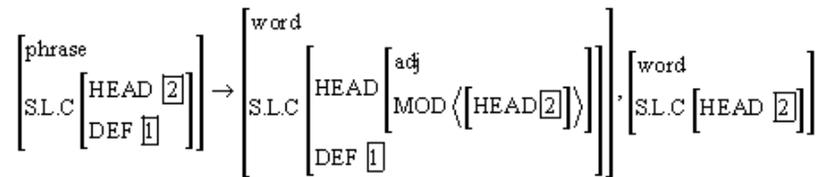


Figure 3. Revised HEAD-MODIFIER SCHEMA for Amharic

Notice that the head value of the head-daughter is re-entered with the head-value of the mother node, thus instantiating the Head-Feature-Principle, however, morphosyntactic definiteness, captured by the feature DEF, is promoted from the adjective to the phrase as a whole. In Figure 3 we have kept the DEF value of the adjective underspecified, capturing the fact that it is possible for an adjective which is either a [DEF +] or a [DEF -] to combine with a noun which is [DEF-]. For example, let us take *t□lk' bet* ‘a big house’ and *t□lk'u bet* ‘the big house’. In both cases the noun *bet* is [DEF -], that is morphosyntactically not marked for definiteness, as it doesn’t attach the definite article. Crucially, we constrain the definite adjective word, that is an adjective which is suffixed with *-u/-w*, to only combine with nominal words or phrases that are marked as [DEF -].

Left-edge marking of definiteness involving quantifiers and cardinals, follow a pattern parallel to the one shown here for the definiteness marking of modifiers.

We now would like to turn to definiteness marking in modified possessive noun-phrases. As outlined in section 1, possessive noun phrases can be derived by means of a possessive suffix. The possessive lexeme-to-word rule given in Figure 4 captures this process:

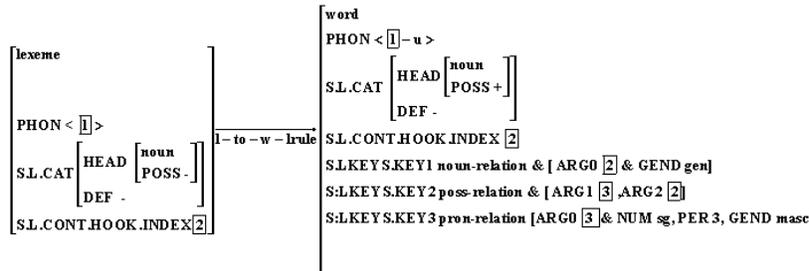


Figure 4. Possessive-lexeme-to-word rule for Amharic

The boolean head-feature POSS identifies the noun word syntactically as possessive. Notice that possessive noun words are syntactically marked as DEF -. The word *bet-u* ‘the/his house’ is thus truly ambiguous—also in its morphosyntactic rendering, with either the specification [HEAD.POSS -] as a definite noun word, or as a possessive 3per.sg.masc-noun-word with the specification [HEAD.POSS +]. In Figure 4, we have included the semantic specifications of a possessive noun. A possessive introduces next to the noun-relation a poss(essive) relation and an abstract pronoun relation representing the possessor. The *poss-relation* takes the possessor as its logical subject and the possessed as its logical object. Not illustrated here, the possessive noun index will be bound by a *definite q_rel* which ensures that the possessive NP is semantically identified as a definite expression. The semantic formalism chosen here is Minimal Recursion Semantics (the reader is referred to Copestake et. al 2004 for more information).

Let us now turn to the modification of possessive nouns. As we observed in Figure 4, the output of the possessive rule is a noun word with a [DEF -] value. As outlined in Section 1, when a possessive noun is modified or quantified, the left-hand daughter of the complex noun phrase necessarily needs to be marked as definite. In Figure 5 we show a possessive noun modified by an adjective:

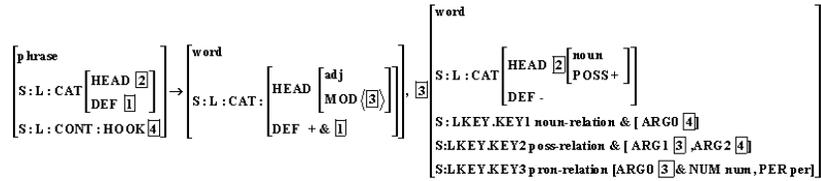


Figure 5. HEAD-MOD-SCHEMA for possessive nouns in Amharic

Again, it is the left-hand daughter, which has undergone the definite lexeme-to-word rule that promotes its morphosyntactic definiteness up to the mother node. It then selects nouns which are specified as DEF- , and, in case of modified possessive nouns as POSS +. We thus exclude the double marking of definiteness in a possessive NP and identify correctly the suffix of the modifier as marker of definiteness while the morphologically identical marker on the noun is rendered as possessive.

5 Conclusion

We have described the distribution of the definite article suffix in Amharic. Using HPSG, we have focused on the HEAD-MODIFIER SCHEMA in our account of the left-edge effect of definiteness marking in Amharic. We have modelled the dependency between morphological nominal possessives and definiteness marking. In promoting definiteness from the left-hand, non-head daughter we have suggested a modular approach to the HEAD-FEATURE PRINCIPLE in order to account for the Amharic facts. In modified possessive nouns we have correctly identified the suffix on the modifier as the definite marker while the suffix on the head noun is identified as the suffixed possessive. Cross-linguistically nominal phrases seem to be double-headed structures, and languages seem to differ widely in where they express morphological agreement, either choosing the head or the specifier as a host. Allowing the ‘collection’ of agreement features from both daughters in a complex NP might be a necessary move also for other languages than Amharic.

References

- Copestake A., D. Flickinger, I. Sag, and C. Pollard. 2004. *Minimal Recursion Semantics*. Manuscript, forthcoming.

- den Dikken, M. 2004. *Definiteness and Agreement in the Amharic Complex Noun Phrase*. Manuscript, The Graduate Center of The City University of New York
- Hellan, L. and D. Beermann, 2005. The 'Specifier' in an HPSG Grammar Implementation of Norwegian. *Proceedings of the 15th Nordic Conference of Computational Linguistics Conference*, ed. S. Werner, 57-64. Joensuu, Finland: University of Joensuu.
- Leslau, W. 1995. *Reference Grammar of Amharic*. Wiesbaden, Germany: Harrassowitz Verlag.
- Lyons, C. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Pollard, C. J. and I. A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.
- Sag, I. A., T. Wasow, and E. M. Bender. 2003. *Syntactic Theory: A formal Introduction*. Number 152 in CSLI Lecture Notes. Stanford: CSLI Publications.
- Wintner, S. 2000. *Definiteness in the Hebrew Noun Phrase*. Manuscript, Institute for Research in Cognitive Science, University of Pennsylvania.

How Universal Is the Pragmatic Detachability Scale? Evidence from Texas German Discourse Markers

HANS C. BOAS AND HUNTER WEILBACHER

1 Introduction

Over the past two decades, discourse markers (DMs) have attracted considerable attention because of their role in marking and negotiating speaker roles (cf. Schiffrin 1985, Brinton 1990, Jucker 1993, Schourup 1999, Fuller 2003, among others). More recently, a number of studies have analyzed DMs in language contact situations in order to determine what types of DMs are borrowed and why (cf. Brody 1987, Maschler 1998, Serra 1998). Depending on the intensity and length of contact between two languages, only selected DMs are borrowed. For example, Gumperz and Hernández-Chaves (1971: 319) show that Spanish-English bilinguals often use English DMs such as *you know* in Spanish. These can be regarded as ethnic identity markers much in the same way that Yiddish expressions like *oi gewalt* characterize the in-group English speech style of some Americans of Jewish heritage.

In other cases, entire discourse-marking systems can be borrowed from one language into another, often fulfilling specific emblematic, pragmatic, or semantic purposes. For example, American German dialects have borrowed English DMs due to intense contact over long periods of time (see Salmons 1990, Fuller 2001). One such DM is *well*, which serves the

Texas Linguistics Society 9: Morphosyntax of Underrepresented Languages

Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, and Jessica White (vol. eds.) and Stephen Wechsler (series ed.).

Copyright © 2007, CSLI Publications.

pragmatic function of framing the content of an utterance in the discourse of German-American bilinguals (see Fuller 2001). According to Matras' (1998) pragmatic detachability hierarchy, DMs such as *well* are borrowed particularly easily because they do not have semantic/pragmatic equivalents in the other language (cf. Fuller 2001 and de Rooij 2000).

In this paper, we compare previous research on English DMs in three American German dialects with more recent data on DMs extracted from a large electronic corpus of spoken Texas German (henceforth TxG). The focus of our investigation is the syntactic, semantic, and pragmatic distribution of *you know* and its German equivalents *weisst du/weisst(e)*. The main questions addressed in this paper are the following: (1) Does *you know* have the same distribution in modern-day TxG as in Pennsylvania German and Indiana German? (2) Do TxG speakers prefer the use of *you know* over its German counterparts *weisst du/weisst(e)*? (3) Has modern-day TxG adopted an English-based DM system similar to other American German dialects? (4) How far do the TxG data support Matras' (1998) pragmatic detachability hierarchy?

The paper is structured as follows: Part one reviews some relevant aspects of English *you know*. Part two summarizes previous research on DMs in American German dialects. Part three discusses the distribution of *you know* in TxG and investigates how it differs from that of other American German dialects. In addition, we examine whether TxG *you know* exhibits the types of properties predicted by Matras' (1998) pragmatic detachability hierarchy. Finally, we will attempt to determine whether TxG is in the process of adopting an English-origin discourse marking system that replaces the German system of DMs.

2 Properties of English *You Know*

Östman (1981) distinguishes several functions of *you know*. The first function is 'face-saving and politeness', which performs an implicit anchoring function such as expressing relative certainty about the acceptability of the subject matter of an utterance, the speaker's emotional stake in the subject matter being breeched (1981: 6-7). According to Östman, *you know* can also lessen the aggressiveness or rudeness of an utterance, as is illustrated by sentences such as *You know, ma'am, if you weren't a lady, I'd punch you right in your face* (Östman 1981: 20).

The second function of *you know* is to mark the beginning or end of a speaker's speech-turn. Implicit anchorage is used by Östman to mean a pragmatic device for tags and hedges which he defines as 'to delimit the area of interest, and to capture the two most basic features of this class of linguistic phenomena' (Östman 1981: 5). These devices can be used to

mitigate the speaker's responsibility for the subject matter of an utterance and to soften the directness of a statement, sometimes implying through their use that in power-semantic terms, the speaker is inferior to the addressee because of his overt implication (*you know*) that the addressee has more information than he has himself. In this sense, expressing politeness is the main purpose of employing *you know* in this way.

Other functions of *you know* include planning and indirectness, i.e. elements of conversation which incorporate aspects of hesitation, repetition, false-starts, and afterthoughts, according to Östman. For example, pauses in speech that are not clearly identifiable as temporary can sometimes be erroneously understood as being relevant markers of transition, as is illustrated by *And then we used to – and I'm an artist you know we used to...* (Östman 1981: 19). These functions do not have to be intentional information-suppression on the part of the speaker, but could instead be indicative of the speaker's uncertainty of what he wants to say. In this way the speaker has failed to plan out his utterance and uses *you know* as a pause-filler. Östman (1981) claims that *you know* has also been characterized in terms of interjections, hedges, and turn-taking devices; in addition, it could be classified in the same category as modal or adverbial particles. This DM's general pragmatic purpose is that of a lexical hedge as it functions to qualify the whole speech act it appears with without actually adding any meaning itself (Östman 1981: 16, 31). As such, *you know* is a lexical-semantic plea for cooperation on the part of the speaker so that the addressee will presuppose the tenability of what he is saying. Moreover, *you know* can be used in the same way as *well* to acknowledge the insufficiency of a response (A: *What time is it?* B: *Well, you know, the sun just came up.* (Östman 1981: 32)). With this brief review of some of the most important functions of *you know*, we now turn to the properties of its German counterpart *weisst du/weisst(e)*.¹

3 Previous Analyses

3.1 English DMs in Texas German and Indiana German

Salmons (1990) analyzes the distribution of English and German DMs in Gillespie County Texas German and Dubois County Indiana German.² He

¹ For different intonation patterns of *you know* in utterance-initial and utterance-final positions, see Östman (1981: 22-24) for a description of the declarative and interrogative contours.

² The two dialects differ in that the former is of primarily Nassau-Hessian origin and thus exhibits many similar central German features, which makes it mutually intelligible with contemporary Standard German (Gilbert 1963, 1972). In contrast, the Dubois County speakers all speak different local varieties (see Salmons 1990: 456).

claims that language contact between English and German has resulted in a shared system of discourse marking that is basically the American English system. It includes borrowed English surface forms, whereas the German system shows only traces of itself (1990: 473). For example, all of Salmons' six speakers have largely lost German modal particles such as *ja*, *aber*, *mal*, and *denn* (1990: 464). At the same time, the speakers have acquired a new English-based system 'that performs some of the functions of the modal particles' (1990: 474).

An important feature of such mixed systems is that they show characteristics of borrowing and convergence, but do not fit definitions of code switching (1990: 455). While the loss of German DMs and simultaneous acquisition of most English DMs such as *well* is explained in terms of mutual convergence (1990: 474), Salmons points out that *you know* and *weisst du/weisst(e)* served similar functions before English and German came into contact with each other in Texas and Indiana. This observation leads him to propose that the distribution of the two DMs should not be described in terms of convergence (cf. Weinreich 1959), but rather in terms of 'overlap' as defined by Woolford (1983: 522) (see Salmons 1990: 469). Based on his field recordings, Salmons (1990) identifies three different functions of *you know* in Indiana German and TxG. The following example illustrates the first function, which is to clarify a point or to establish a shared knowledge with the hearer.

(1) Ich war hier mehre Jahre zurick, waren an Reise gewese mit die
Feed Company, you know, womit ich arbeite due ...

'I was here back several years ago, [we] were on a trip with the feed
company, you know, that I work for ...' (Salmons 1990: 460)

In (1), the speaker employs *you know* while telling a story about traveling for the feed company. This use signals to the hearer that the speaker intends to clarify whether the hearer has understood the first statement of his sentence. Alternatively, *you know* in this context could signal that the speaker and hearer share the same background knowledge (see Schifffrin (1987: 267-68), Salmons (1990: 459)). The second function of *you know* is that of hesitation or introspective use (see Schourup (1985: 102) and Salmons (1990: 459-60)), as illustrated by the following sentence:

- (2) Das is nicht schlim; das is alright; das is bloss, da sitzt man in Haus, man kann nichts doan, you know, und alles, muss da bei die Phone sitzen.

‘It’s not bad, it’s all right, it’s just that, there you sit in the house, you can’t do anything, you know, and all, [I] have to sit by the phone.’ (Salmons 1990: 460)

Another function identified by Salmons is that of self-repair, as discussed by Östman (1981: 29-31) employed here to smooth over a false start:

- (3) Aber das hat auch mit de ... you know, wenn sie mit in die Schul gehn...

‘But that has to do with the ... you know, when they go along to school ...’ (Salmons 1990: 460)

With this brief review of Salmons’ (1990) classification of *you know* in Texas German and Indiana German, we now turn to Fuller’s (2001) analysis of the distribution of *you know* in Pennsylvania German, which also considers the distribution of its German counterpart *weescht*.

3.2 Pragmatic Detachability of DMs in Pennsylvania German

Based on a total of thirty hours of conversational data collected among 18 speakers, Fuller (2001) discusses Pennsylvania German DMs in three categories: (1) DMs that have counterparts in both languages, (2) English DMs that do not have semantic/pragmatic equivalents in German, and (3) German DMs that do not have semantic/pragmatic equivalents in English. DMs of each category are listed in Table 1. Fuller’s (2001) analysis crucially depends on Matras’ (1998) account of grammatical borrowing, which we will now discuss briefly. Matras suggests that in language contact situations, the donor language is often pragmatically dominant, which can lead to mixed DM systems (DMs of donor and recipient languages converge), or the complete borrowing of the DM system of the donor language. Matras (1998) establishes three scenarios for DM borrowing: (1) the donor language is pragmatically dominant; (2) the change leading to convergence is not only gradual but also gradational; (3) there is a hierarchy of pragmatic detachability, and those DMs that are at the top of the hierarchy (i.e., pragmatically more detachable) will be borrowed first. Whether a DM is easily borrowed depends on three different scales.

Discourse marker	<i>N</i>	Discourse function
<i>well</i>	106	indicates that the subsequent utterance may not be what is expected by the hearer
<i>so</i>	193	links two utterances or assumptions causally
<i>y'know</i>	215	emphasis/focus; presents information as shared, creating common ground
<i>weescht</i>	154	emphasis/focus; presents information as shared, creating common ground
<i>but</i>	224	contrastive conjunction
<i>aber</i>	201	contrastive conjunction
<i>ja</i>	1	contrast an emphasis
<i>mal</i>	15	indicates the limited duration of the action referred to in the utterance

Table 1. *DMs in Pennsylvania German and their functions* (see Fuller 2001: 356)³

The first scale is the pragmatic detachability scale. Elements that organize the speech event are perceived as gesturelike,⁴ situation-bound devices and are therefore detachable from the content message of the utterance, according to Matras (1998: 309). Such elements are more turn-related (e.g. *well*) and are borrowed before more content-related ones such as *but* (which functions to contrast the meanings of clauses). The second scale is the category-sensitive scale, which illustrates that elements that are more lexical or deictic are borrowed last. In other words, elements that are not easily analyzed in terms of lexical meaning (e.g., *well*) are predicted to be borrowed more easily than more highly lexical items (e.g. *you know*). Matras' (1998) third scale, the semantic scale, is used to measure DMs that identify contrast, restriction, or change, which are more easily borrowed than those that mark addition, elaboration, or continuation. Therefore,

³ Fuller's (2001: 356) original table appears to have switched the order of *y'know* and *but*. Whereas her table shows *y'know* followed by *but*, *weescht*, and *aber*, we have corrected the order of DMs in our table to what we believe was the intended order of DMs in Fuller's original table.

⁴ Gesturelike: In the area of conjunction and focus particles, it is hypothesized that gesturelike properties go together with the semantics-pragmatics of contrast, change, and restriction. In the domain of sentence particles, hesitation markers, fillers and tags, it can be assumed that the less lexical content an expression has and the less analyzable it is to the speaker, the more gesturelike and situation-bound it is likely to be (Matras 1998: 310).

contrastive *but* is more likely to be borrowed than the additive conjunction *and*.

Following Matras' (1998) proposals, Fuller (2001) splits Pennsylvania German DMs into three distinct groups, each differing in their level of pragmatic detachability. Members of the first group, which consists of *well* and *so*, signal a change in speakers, and their content is not easily analyzable, because they are nonlexical (Fuller 2001: 360). This combination of factors makes these DMs highly pragmatically detachable, which explains why they have been easily borrowed into PG from English (a fact that is also supported by the frequency data), according to Fuller.

Her second group, which includes *you know* and *but*, are ranked lower on the pragmatic detachability scale and are used in variation with their German counterparts *weescht* and *aber*. She points out that the English DMs are not borrowed that easily, because they are lexical. Although *you know* exhibits content-related functions, it is also frequently used in a turn-related fashion, which places it high enough on the pragmatic detachability scale to be borrowed into PG. With respect to the frequency of English DMs and their German counterparts Fuller claims that *you know* 'has operational features that contribute to its borrowability, and it is thus present at a high frequency: 215/369, or 58 percent, of the contributed *y'know weescht* tokens are *y'know* in these PG data' (2001: 361). Another important fact pointed out by Fuller is that both *y'know* and *weescht* appear in the same discourse contexts, that is, either at the beginning, in the middle, or at the end of an utterance (2001: 362-363). Besides its literal usages indicating mutual knowledge (see also Östman 1981, Holmes 1986, Brinton 1990), a bid for common ground, a confirmation check, and an emphasis to draw attention to a point, among others, *y'know* and *weescht* can also be used nonliterally. An example is a case where information is clearly not known to the hearer, where both DMs can be used to indicate that an utterance contains new information and opinions (2001: 362). These observations lead her to the conclusion that there is 'a gradual turnover from the recipient-language discourse-marking system used in the donor language' (2001: 363).

The third group, which includes *ja* and *mal*, are at the lowest level of Fuller's hierarchy, because 'they are not pragmatically detachable, and they do not have English equivalents that could replace them' (2001: 365). Since they occur only infrequently in PG, Fuller proposes that they signal a loss of German DMs, or, in other words, that "the English discourse-marking system is taking over the language" (2001: 366).

The distribution of English and German DMs lead Fuller to propose that Pennsylvania German exhibits a gradual and gradational turnover from the recipient-language discourse-marking system to the DM system of the

donor language (i.e., English), where the motivation for borrowing is based on pragmatic functions (2001: 353). Thus, DMs that are high on the detachability scale (e.g., *well*) are borrowed more easily than those that are lower (e.g., *you know*), as illustrated by Table 1 (cf. Fuller 2001: 363). At the same time, German-origin DMs that persist are pragmatically detachable vestiges. She concludes that the likelihood of a DM to be lost or borrowed is predictable based on Matras' (1998) pragmatic detachability scale (Fuller 2001: 367). Having surveyed the distribution of *you know* and its German counterparts in three American German dialects, we now turn to data from present-day Texas German. We begin with a brief discussion of the present situation of TxG, and then we address issues of data collection and analysis.

4 Distribution of *You Know* in Present-Day Texas German

At the beginning of the 21st century, only an estimated 8-10,000 fluent speakers of fifth and sixth generation Texas Germans remain. The number of semifluent speakers using a drastically reduced inventory of phrases based on the language of their forbearers is estimated to be 4-6,000. Since the great majority of the remaining fluent and semifluent speakers of TxG are sixty years and older and the dialect is not learned by the younger generations, it is considered a critically endangered dialect. This sharp decrease in speakers puts TxG on the list of about 3,000 languages and dialects world-wide that are expected to go extinct by the end of the 21st century (see Boas 2005: 1031).

4.1 Data Collection and Analysis

In 2001, the Texas German Dialect Project (TGDP) was founded at the University of Texas at Austin in order to record and archive as many remaining TxG speakers as possible. Since 2002, members of the TGDP have interviewed more than 180 fluent speakers of TxG in New Braunfels, Fredericksburg, Doss, Brenham, Victoria, Houston, Schulenburg, Black Jack, Spring Branch, Bulverde, and Crawford, among other locations. The interviews have resulted in the recording of different types of data, including open-ended sociolinguistic interviews and elicitation tasks, as well as written demographic questionnaires (see Boas 2003). The audio recordings of the open-ended interviews are transcribed and translated using the ELAN annotation software, and subsequently stored in the Texas German Dialect Archive (see Boas 2006), which is accessible over the

internet.⁵ The resulting electronic corpus is in XML-compatible format and is searchable using a web-based concordancer interface that employs PERL regular expressions (Hall and Schwartz 1997).

In October 2005, we accessed the on-line archive with the concordancer interface, searching for the distribution of *you know* and its German counterparts in the transcriptions of the informants' speech. We searched the transcripts of open-ended interviews with 70 informants (each interview is about 50-55 minutes in length), which were conducted between February 2002 and September 2005. The electronic corpus contains a total of 305,429 German words uttered by our informants.⁶ Our search resulted in a list of 599 instances of *you know* and 5 instances of *weisst du/weisst(e)* used as DMs (as opposed to their use in questions (e.g., *You know where the store is?*) or other types of grammatical constructions). The list was then divided into five different categories, according to the functions of the DMs: awareness of shared knowledge, clarification of common knowledge, indication of hesitation, self-repair, and appeal for understanding. In the following sections, we first discuss the distribution of the English and German DMs for each of the categories, and then we compare our findings with the results of Salmons (1990) and Fuller (2001).

4.2 Awareness of Shared Knowledge

The most common use of *you know* and *weisst du/ weissst(e)* in present-day TxG is to encourage the hearer to accept what the speaker is saying as shared knowledge (cf. Östman 1981, Schiffrin 1987). For example, in (4) the speaker aims to convince the hearer to recognize that there was only one person among a group of people who became a teacher. This sentence illustrates that *you know* is often used to introduce new knowledge, where *you know* signals the beginning of new information (cf. Brinton 1990):

- (4) Und der einzige Lehrer, die anderen warn alle you know Haus
and the only teacher the others were all you know house
frauen geworden oder Farmers oder was.
wives became or farmers or something

‘And the only teacher, the others all became you know housewives

⁵ The URL of the TGDP is <http://www.tgdp.org>. For more information on ELAN, which was developed by the Max Planck Institute for Psycholinguistics at Nijmegen, see <http://www.mpi.nl/tools/elan.html>.

⁶ The electronic corpus is enlarged constantly as more recordings are transcribed, translated, and stored in the online archive.

- or farmers or something (else).’ (1-25-1-3-a)⁷
 (5) Mit de dope da you know drugs.
 ‘With the dope there, you know drugs.’ (1-98-1-130-a)

Example (5) shows that *you know* is also employed to introduce new information at the end of sentences. In this case, the speaker introduces the borrowed word *drugs* in combination with *you know* in order to make sure that the hearer understands what is meant. In (6) the speaker answers a question about the class size of elementary school classes in a small rural community northwest of Fredericksburg in the late 1940s. After acknowledging that there were quite a few students, the speaker uses *you know* to encourage the hearer to acknowledge the actual class size as being 45 or 50 students. This use of *you know* is an instance of what He and Lindsey (1998) label an ‘information status enhancing device’, used to highlight a specific point (in this case, the more exact number of students).

- (6) Oh da waren mehr divan you know finfundverzig fuchzig.
 Oh there were more of that you know five-and-forty fifty
 ‘Oh, there were more of those, you know, forty-five, fifty.’
 (1-54-1-2-a)

Similar to Fuller’s (2001) description of Pennsylvania German, our data show that *you know* and *weisst du/weisst(e)* can occur in the same contexts in TxG. The following example is a response to a question about who immigrated to Texas in the 19th century.

- (7) Unser Verwandte weisste und da.
 our relative you know and there.
 ‘Our relative you know and there.’ (1-28-1-22-a)

So far, we have shown that this particular use of the two DMs in modern-day TxG is similar to that described by Salmons (1990) for (older) TxG and Indiana German, and by Fuller (2001) for Pennsylvania German. However, there is one important difference between Fuller’s (2001) data set and ours. Although the English and German DMs can occur in the same contexts in

⁷ The combination of numbers at the end of sentences point to files in the Texas German Dialect Archive (<http://www.tgdp.org>) that contain them. The first number refers to the interviewer, the second number identifies the informant, the third number represents the interview with the informant, and the fourth number indicates the section of the interview. The designations ‘a’ and ‘v’ refer to audio and video, respectively.

both Pennsylvania German and TxG, they differ drastically in their numerical distribution. When used to indicate awareness of shared knowledge, we found 539 tokens of *you know* (99.63 percent), but only two tokens of *weisst du/weisst(e)* (0.37 percent) in the TxG corpus. This is in stark contrast to Fuller's data set, where 58 percent of the overall number of *y'know weescht* tokens are *y'know* (2001: 361). Since Fuller does not include specific frequency counts when discussing the different uses of the two DMs, we will return to this point after discussing the other functions of the two DMs (and their frequencies) in TxG. Our discussion of the frequency data will also shed light on the question of whether Matras' (1998) pragmatic detachability hierarchy makes the right predictions about the distribution of bilingual DMs in TxG.

4.3 Clarification of Common Knowledge

The second major function of *you know* in TxG is to mark knowledge that is already known to the speaker. This function, which is similar to its Pennsylvania German counterpart labeled by Fuller (2001: 361) as 'confirmation check', differs from the one above in that it does not introduce new information or attempts to encourage the speaker to accept the information following *you know* as new information. Consider (8), which comes at the end of a lengthy narrative about World War I explaining how Germany had lost the war. Here *you know* serves to repeat common knowledge in order to make sure that the hearer knows about the content of the utterance. This strategy re-affirms a mutual understanding about the situation.

- (8) In neunzehnzwanzig Deutschland war nichts, you know.
 'In 1920 Germany was nothing, you know.' (1-63-1-2-a)
- (9) Und denn du kocht das icing you know iber und tue das iber.
 and then you cook that icing you know over and put it over
 'And then you cook the icing, you know, over and put it over [it].'
 (1-60-1-15-a)
- (10) Aber die hat ... you know gutes Gemis von Leut gekauft.
 but she has ... you know good vegetables from people bought
 'But she has, you know, bought good vegetables from people.'
 (1-39-1-11-a)

A similar state of affairs can be observed in (9), where the speaker uses *you know* to indicate that both she and the interviewer are aware of (a) what

icing is, and (b) that there is a pre-arranged set of steps for preparing cakes, i.e. that cooking the icing and putting it on the cake is the logical next step after the cake has come out of the oven and has cooled off. Similarly, in (10) the speaker uses *you know* to reference a previous statement that she made about her relative being a good cook. In this case, *you know* indicates that the validity of her statement will already be anticipated by the hearer. These uses of *you know* are identified by Schourup (1985) as one its central functions:

Y[ou] k[now] indicates that the speaker expects that there is no communicatively significant discrepancy between what is now in the private world and what is now in the other world, with respect to what is now in the shared world.
(Schourup 1985: 102)

Similar occurrences are found with the German DM *weisste* as the following sentence shows:

(11) Und die - die Männer, weisste, was auch durch die Schule
And the - the men, you know, which also through the school
gegangen warn un was.
went gone and what

‘And the men, you know, who also completed school and so on.’
(1-28-1-23-a)

The example in (11) is part of a conversation about attending a vocational school in San Antonio. Before uttering (11), the informant is talking about the fact that all teachers had to go through the school in order to get an education. The use of *weisste* reiterates that the teachers attending the school also included male teachers. Comparing the frequency of the DMs with this function, we find a total of 22 instances of *you know* and only 2 instances of *weisst du/weisst(e)*.

4.4 Indication of Hesitation

You know can also be used as an indication of hesitation on the part of the speaker when the subject matter is difficult or evocative of strong emotions. This often is the case when a speaker has begun to discuss something emotional, like the loss of a friend or relative, or something that could be

embarrassing, such as a relative's alcoholism. Such instances are often precipitated by pauses as the following examples illustrate.

- (12) Andere da ... sind gestorben oder ... oder you know, das is das,
 others there have died or or you know that is that

wenn's einfach so allmählich komm.
 when-it simply so slowly comes

'Others there died, or, you know, that is it when it simply goes like that.' (1-25-1-21-a)

- (13) Alle die – da war so 'n Art uh you know Krieg gewesen da
 all those there was so a sort uh you know war was there

sehr ... uh unruh
 very uh unrest

'All those- there was a sort of uh, you know, war there and a lot of uh unrest.' (1-59-1-4-a)

This use of *you know* is similar to Salmons' (1990: 458) discussion of Schourup's (1985) notion of introspective usage of DMs in similar situations. This appears to be congruent with our example of hesitation in (12), where *you know* is not merely a pause-filler, though it does occur in conjunction with pauses and hesitation. The corpus contains a total of four instances of this function of *you know*, but no comparable examples with *weisst du/weisst(e)*.

4.5 Self-Repair

Salmons (1990: 460) points out that sometimes 'self-repair is set up with *you know*'. In our data, we found nine examples where a speaker misspeaks and *you know* is used to mark the mistake in a way that expresses something like *you know what I meant to say* (we did not find any instances of *weisst du/weisst(e)* in this context). Consider (14), where the speaker says the wrong number and uses *you know* to segue into the correction when talking about the end of World War I:

- (14) Uh es nähmte denn bis acht you know neunzehnachtzehn wenn
uh it took then until eight you know nineteen eighteen when

der Krieg vorbei war ...
the war over was ...

‘It took then until eight, you know, nineteen eighteen when the
war was over...’ (1-63-1-2-a)

- (15) Es wächst nach San Anton ganz wei-, you know, ganz viel.
it grows to San Antonio very fa-, you know, very much
‘It grows to San Antonio very fa-, you know, a whole lot.

(1-98-1-13-a)

Similarly, when informant 98 talks about urban sprawl in the San Antonio area in (15) she stops short of completing the phrase *ganz weit* (‘very far’), interrupts her sentence with *you know*, and then completes it with *ganz viel* (‘very much’). Sentences (16) and (17) provide additional examples of *you know* being used for self-repair.

- (16) Denn ge- denn das Blut you know - UH Messer genom un
then Pr then the blood you know - uh knife taken and

in uh in Hals hier
in uh in throat here

‘Then the blood you know – I took the knife and put it in the throat
here.’ (1-7-1-21-a)

- (17) Mein Grossvater war war verheiradet bevo- you know ... before
my grandfather was was married befo- you know ... before

uh uh anyhow uh – denn he married.
uh uh anyhow uh – then he married

‘My grandfather was married before, you now, before, uh, anyhow
uh, then he married.’ (1-92-1-8-a)

4.6. Appeal for Understanding

You know can also be built into a statement as an appeal for understanding or empathizing with the speaker's disposition toward the information contained in the statement. This use of *you know* is different from the simple use signaling awareness of knowledge discussed in 4.2 in that it includes an emotional component which reveals the speakers' feelings. Consider example (18), which is uttered during a conversation about activities during lunch breaks at school in the 1930s.

- (18) You know ich kann mich gar nicht mehr erinnern.
 you know I can myself totally not more remember
 'You know, I totally can't remember that myself.' (1-8-1-4-a)

Before uttering (18), the informant talks about playing baseball during breaks while in sixth and seventh grade. However, when asked about what children used to play during lunch breaks in elementary school, she admits that she cannot remember what games she used to play when she was that young. The use of *you know* at the beginning of this sentence thus suggests that the informant may feel slightly embarrassed, which explains her use of the DM. That is, she intends to appeal for the understanding of the hearer.

Sometimes, *you know* is also used to mark a sentence as contradictory to what one might expect. In this case, the DM is used to appeal to the hearer's attention that the utterance may contain some unexpected information that the hearer was unaware of previously. An example of this use of *you know* is (19), where the informant explains why she did not understand the reasons for not being allowed to speak German in school:

- (19) You know ich war ... uh ... uhm ... acht I guess.
 'You know I was ... uh ... uhm ... eight I guess.' (1-82-1-8-a)

Sentences (20) - (21) are other examples illustrating this particular use of *you know*. (20) is uttered in a context where the informant expresses her remorse that she and her husband did not teach their kids German when they were growing up.

- (20) You know das ist jetzt unser ganze Schuld.
 you know that is now our whole guilt
 'You know, we're really guilty of that.' (1-85-1-8-a)

- (21) You know die die darf nicht nach die Schule gehn.
 you know they they allowed not to the school go
 ‘You know, they were not allowed to go to school.’ (1-59-1-12-a)

Before uttering (21), informant 59 talks about school education in the early 20th century, pointing out that school education was generally available for every child, even in remote areas of the Hill Country. However, (21) reminds the hearer that there was an (from today’s point of view unfair) exception to the availability of school education beyond the fifth grade, namely, that girls were not allowed to continue their education once they had reached a certain age. As with the previous usages of *you know* and *weisst du/weisst(e)* can appear in the same pragmatic contexts. However, the distribution of the DMs is extremely unbalanced, with *you know* appearing 25 times in our corpus, and *weisst du/weisst(e)* only once, as shown in (22):

- (22) Weisst, er hat un- er hat un- ... ein Geschäft gehabt.
 you know he has un- he has un- a store had
 ‘You know he had a store.’ (1-1-1-29-a)

4.7 Comparison with Salmons’ (1990) and Fuller’s (2001) Data

Table 2 summarizes our discussion of the present-day TxG data in the sections above. Comparing our data with Salmons’ (1990) and Fuller’s (2001) classifications of the functions of *you know* and *weisst du/weisst(e)* we note a number of important similarities. *You know* is most frequently used in our data to indicate that both the speaker and the addressee are aware of shared knowledge. This suggests that for Texas Germans establishing a communicative base-line may be a very important element of discourse, a point already made by Fuller (2001).

Our data are also similar to Salmons’ (1990) and Fuller’s (2001) accounts in that they reflect the use of *you know* to indicate the clarification of common knowledge, hesitation, and self-repair. In concordance with Salmons (1990) and Östman (1981), we also find that the DMs appear most frequently in the middle of utterances, as Table 3 illustrates. This distribution can be partially attributed to the fact that the DMs often occur in conjunction with pauses and between clauses.

Category	<i>you know</i>	<i>weisst du/weisst(e)</i>
Aware of knowledge shared	539	2
Clarification of common knowledge	22	2
Indication of hesitation	1	0
Self-repair	12	0
Appeal for understanding	25	1

Table 2. *Summary of pragmatic contexts in present-day TxG*

Position of DM	<i>you know</i>	<i>weisst du/weisst(e)</i>
Utterance-initially	82	1
Utterance-medially	359	4
Utterance-finally	158	0

Table 3. *Distribution of discourse markers*

Interestingly, there are also some differences between our corpus data and the data reported by Salmons (1990) and Fuller (2001). For example, Salmons claims that ‘the slots where English DMs occur are not natural slots in German to contemporary European speakers’ (1990: 362-363). Based on informal interviews with six native German speakers residing in Europe, we could not confirm this type of restriction for our data extracted from the TxG corpus. That is, given a range of options and contexts, native German speakers are indeed capable of filling *weisst du/weisst(e)* into empty slots of sentences where *you know* has previously been deleted. For example, after erasing *you know* from a sentence such as *You know das ist jetzt unser ganze Schuld* (1-85-1-8-a), we asked our European German informants what types of DMs they could fill in, given different contexts. Answers to this, as well as fifty other examples randomly chosen from our dataset, included *ja*, *doch*, *naja*, and *weisst du/ weisst(e)*, depending on the context. The results from this preliminary experiment suggests that *you know* and *weisst du/weisst(e)* may be used interchangeably in TxG. Additional evidence comes from two sources. First, we created a list of twenty-five English sentences including *you know* in various contexts. In January 2006, we asked ten Texas German informants from Fredericksburg and Doss to translate the sentences from English into TxG. Except for three

mistranslations, all informants successfully translated *you know* as *weisst du/weisst(e)* into TxG. The second piece of evidence for the interchangeability of the two DMs comes from Fuller's Pennsylvania German data, which demonstrate that 'the placement options for these DMs are the same (2001: 361-62).

Another difference between present-day TxG and the data presented by Salmons (1990) and Fuller (2001) is the occurrence of *you know* in contexts that have previously not been described in detail. The first type concerns the use of *you know* preceding or following borrowed words and phrases as illustrated by the following examples.

- (23) Was ist das Wort wo die punished 'em you know.
 what is the word where they punished them you know
 'What is the word, they punished them, you know.' (1-1-1-6-a)
- (24) Un you know my dad done often said neunzehnneuzehn das war...
 'And you know my dad done often said nineteen nineteen that
 was...'. (1-21-1-9-a)
- (25) Oh der war raus und hat tires gechanged, you know.
 oh it was out and had tires changed, you know
 'Oh he was outside and changed tires, you know.' (1-114-1-8-a)
- (26) Die ham ein von die was did hat you know foreign language.
 they have one of the was did have you know foreign language
 'They have ... one of them knew, you know, a foreign language.'
 (1-78-1-10-a)
- (27) Well, das musste alles you know with elevation geschossen
 well that must everything you know with elevation shot
 werden.
 been
 'Well, all of it, you know, had to be shot with elevation.'
 (1-90-1-15-a)

The uses of *you know* in (23) – (27) signal shared knowledge or an attempt to appeal to the hearer to accept what the speaker is saying. Technically, this use could be considered a subtype of the more general type of 'confirmation check' described in section 4.3. We attribute the high frequency of this particular use of the DM to the fact that all interviewers

speak Standard German with the informants, who are often aware of the linguistic differences. In other words, the differences between the two varieties – including the misperception among some speakers of TxG that their variety is somehow inferior vis-à-vis Standard German – and the fact that the interviewers also speak English fluently may lead the TxG speakers to employ *you know* in the context of a word or phrase that is borrowed from English.

Another phenomenon that we have found in our corpus data, and which is not explicitly mentioned by Salmons (1990) or Fuller (2001), is the combined use of two DMs next to each other. Consider (28), where *you know* is followed by *well* in the middle of the sentence.

(28) Und denn hier sind die Kinder, die sind alle, you know, well,
and then here are the children they are all you know well

Englisch und die Eltern, sie waren Deutsch.
English and the parents they were German

‘And then here are the children, they are all, you know, well,
English, and the parents, they were German.’ (1-63-1-6-a)

At first sight, it is not clear how to classify the use of *you know* in (28). It could be regarded as indicating hesitation on the part of the speaker who, while uttering (28), is in the process of realizing that there are pronounced linguistic and cultural differences between the older and the younger generations. This analysis is supported by the subsequent use of *well*, which has been classified as a conversational-coherence marker which indicates that the speaker’s utterance may not fit with the presuppositions of the hearer (Schiffrin 1987: 127). This view suggests that *well* may be used to emphasize the meaning of *you know* in this particular context in order to heighten the feeling of uncertainty and embarrassment that overcomes the speaker while uttering (28). Alternatively, the use of *you know* in combination with *well* could be interpreted as an attempt to establish mutual knowledge with the hearer (cf. section 4.2 above). Often it is not possible to clearly distinguish between the different functions of the DMs. For example, the combined use of *you know* and *well* in (29) may at first sight appear to signal self-repair (cf. section 4.5 above).

(29) Wenn wir jetzt Kinder gehabt hätten, you know, well wir haben
if we now children have had you know well we have

sie hat zwei Kinder gehabt.
she has two children had

‘If we would have had children now, you know, well, we have she
had two children.’ (1-94-1-18-a)

In sentence (29) the speaker tells a story about how he and his wife did not have children, only to go on to say how his sister had two children. The use of *you know* in the middle of the sentence could be regarded as an appeal for understanding if the sentence ended naturally by naming the consequences of having children. However, the conditional sentence is interrupted halfway and is continued instead by the regular declarative sentence *sie hat zwei Kinder gehabt* (‘... she had two children’). This suggests that *you know* is employed here both as an appeal for understanding and as an indicator of hesitation followed by *well* to start a new thought. This overlap in functions becomes even more complicated when *you know* is both preceded and followed by *well* as in the following example:

(30) Well die kam’n zu mich, und ich well you know well meine uh ich
well she came then to me and I well you know well my uh I

hab ein Bruder gehabt.
have a brother had

‘Well, she came to me, and I well you know well my uh I had a
brother.’ (1-39-1-15-a)

The frequency of *you know* vis-à-vis *weisst du/weisst(e)* is another point that sets our data apart from those reported by Salmons (1990) and Fuller (2001). In contrast to Salmons, we have already shown above that the two DMs are capable of occurring in the same pragmatic contexts, i.e., they can be substituted for each other. This would lead us to expect that *weisst du/weisst(e)* would also occur relatively frequently in our corpus data. However, this expectation is not borne out. In our data we find only an 0.83 percent occurrence of *weisst(e)/weisst du* with an overwhelming occurrence of 99.17 percent of *you know* compared to 58 percent for *you know* and 42 percent for *weescht* in Fuller’s PG data. In the following section we discuss

a possible explanation for the unequal distribution of the two DMs in present-day TxG.

5 Evidence for Matras' (1998) Pragmatic Detachability Hierarchy?

One way of explaining the low frequency of *weisst du/weisst(e)* in present-day TxG would be to follow Fuller's (2001) proposals, i.e. to adopt Matras' (1998) concept of pragmatic detachability to explain the borrowing of English DMs into TxG. In this view *you know* has been borrowed into TxG because it has a number of operational features, i.e. content-related and turn-related functions, which contribute to its borrowability (Fuller 2001: 361). Following this line of reasoning requires us to consider a number of other important points, to which we now turn.

The first point concerns the present-day status of TxG DMs vis-à-vis Salmons' (1990) account. Concerning the distribution of *you know* and *weisst du/weisst(e)* Salmons claims that 'it is not surprising, then, to find these two surface forms, one English and one German, in variation for many speakers' (1990: 469). Although he does not provide exact frequency counts reflecting the variation between these English and German DMs, it appears as if the distribution is significant enough for Salmons to characterize it as 'overlap' (1990: 469). While this overlap undoubtedly still exists in present-day TxG, our corpus data demonstrate that at the beginning of the 21st century there is virtually no variation left between the two DMs (99.17 percent of the tokens in our corpus are *you know*). Our comparison of present-day TxG with Salmons' data, which was collected in the early 1980s, thus indicates a dramatic loss of *weisst du/weisst(e)* within two decades.⁸ As such, our data are also in stark contrast to the distribution of the two DMs in Pennsylvania German, where *y'know* exhibits a much higher frequency (58 percent of tokens) (Fuller 2001: 361). This divergence in distribution would suggest that present-day TxG *you know* is located at a higher level of Matras' (1998) pragmatic detachability scale than it is in Pennsylvania German. As such, our data would also indicate that the turnover from the recipient-language discourse-marking system to that of the donor language is much further advanced in present-day TxG than previously thought. Although such an explanation sounds intriguing at first

⁸ The difference in variation could also be caused by discrepancies between corpora. The data cited by Salmons are based on fieldwork with three TxG speakers in Gillespie County. In contrast, our corpus data are based on interviews with 70 TxG speakers from various locations across central Texas.

sight, it is imperative that we consider it in the overall context of Matras' scale.

The second point of significance is thus the question of whether other TxG DMs exhibit frequencies that are similar to those reported by Fuller (2001) for Pennsylvania German.⁹ According to Fuller, the contrastive conjunctions *but* and (its German counterpart) *aber*, are roughly at the same level of the pragmatic detachability scale as *you know* and *weisst du/weisst(e)*, with the English DM accounting for 47 percent of all tokens, and its German counterpart for 53 percent (2001: 364). The distribution of these DMs is quite different in the electronic corpus of present-day TxG, where we find 1493 instances of contrastive *aber* (94 percent of all tokens), but only 93 instances (6 percent) of contrastive *but*. This difference in distribution suggests that in TxG, in contrast to Pennsylvania German, *but* and *aber* are not part of a pattern indicating 'borrowing of DMs, possibly as part of a turnover in the discourse-marking system from a German system to an English one' (Fuller 2001: 365). If we were to establish a hierarchy similar to Fuller's (2001) ranking of DMs based on Matras' (1998) pragmatic detachability scale (cf. Table 1), the order for *but/aber* and *you know/weisst(e)/ weisst du* would be reversed.

Next, consider the distribution of German-origin DMs such as *ja*, *mal*, and *doch*, which 'do not have easy English equivalents that could replace them' (Fuller 2001: 365). Their infrequent use in Pennsylvania German is explained by the fact that they are 'fairly low on the pragmatic-detachability scale', which in turn lends further support to Fuller's hypothesis that 'there is a general trend away from the use of German-origin DMs in PG and that an English discourse-marking system is taking over the language' (2001: 366). Similarly, Salmons (1990: 462) claims that such DMs 'have been lost or radically reduced in function and frequency for even the most fluent speakers in central Texas and southwestern Indiana.' However, a search of our corpus data does not confirm Salmons' and Fuller's claims. For example, contrary to Salmons' (1990: 462) assertion that *doch* 'is essentially nonexistent among Texas and Indiana Germans', we find 118 tokens of it in our corpus. Other German-origin DMs that we found in our

⁹ Both *well* and *so* occur at a much higher frequency (796 and 2211 instances, respectively) in our corpus than any other DMs analyzed in this paper. As such, their distributions are similar to those reported by Salmons (1990) and Fuller (2001) vis-à-vis other English DMs. Note, however, that it is problematic to compare the properties and frequencies of *well* and *so* with those of other English DMs that have German equivalents. The former have virtually no 'competition' from pre-existing DMs in the recipient language, whereas the latter are capable of occurring in the same contexts and are thus susceptible to turnover of the type reported by Fuller.

corpus, but which are claimed to have been almost completely abandoned, include *mal* (138 tokens) and *ja* (142 tokens), besides others.¹⁰ The frequency of these German-origin DMs in present-day TxG thus shows that they have not yet been lost completely, but are rather unexpectedly well represented.

6 Conclusions

Our analysis of present-day TxG has demonstrated that *you know* and its German counterpart *weisst du/weisst(e)* occur in similar pragmatic contexts, namely (1) to indicate awareness of shared knowledge, (2) to clarify common knowledge, (3) to indicate hesitation, (4), to appeal for the hearer's understanding, and (5) as a self-repair strategy. Based on data from a large electronic corpus we have shown that the distribution of the two DMs is similar to that reported by Fuller (2001) for Pennsylvania German.

Besides these similarities, we have also shown that our data differ significantly from those reported by Salmons (1990) and Fuller (2001). The first difference concerns the frequency of English and German DMs. For example, we found that TxG *you know* accounts for an overwhelming 99.17 percent of the *you know/weisst(e)* tokens in our corpus, compared to 58 percent *you know* tokens in Fuller's Pennsylvania German data. This high frequency has led us to hypothesize that TxG *you know* may be more pragmatically detachable than its Pennsylvania German counterpart. This, in turn, would suggest that the turnover from the recipient language discourse-marking system to that of the donor language (English) has progressed further in TxG than in Pennsylvania German. However, we have been unable to confirm our hypothesis as we also found frequency data on TxG *but* and *aber* that could be interpreted as counter evidence. More specifically, these DMs do not conform to Fuller's proposal that they are located at the same mid-level position on Matras' (1998) pragmatic detachability scale as *you know* and *weisst du/weisst(e)*. Furthermore, German-origin DMs such *ja* and *mal*, which are supposed to have basically dropped out of use in both TxG and Pennsylvania German appear with a surprisingly high frequency in our corpus despite their location at the very bottom of Matras' pragmatic detachability hierarchy.

¹⁰ Examples of *mal* and *ja* include the following:

- (i) And hat sie gesagt, ich will mal wissen, wie freundlich Neu Braunfels ist.
'And the she said I'd like to know how friendly New Braunfels is.' (1-59-1-18-a)
- (ii) Und denn denn habe ich ja ein hübsches Mädchen gesehen un uh hab getanzt
'And then I saw a beautiful girl and danced.' (1-92-1-11-a)

At this point, our preliminary analysis suggests that Matras' (1998) hierarchy is not universally applicable to language contact situations across the board, because it does not appear to make the correct predictions about the distribution of bilingual DMs in TxG. The divergent properties of DMs in the donor dialects forming the basis for TxG and Pennsylvania German (cf. Raith 1992, Trudgill 2004, Boas 2005) could be one possible explanation for this discrepancy. Another would be to propose that some DMs in TxG are perhaps more resilient to being replaced than others. If this were the case, we would have to develop ways of measuring different synchronic and diachronic factors that influence the turnover of individual DMs. Clearly, further research needs to address the different questions raised in this paper in order to arrive at a better understanding of (1) the distribution of DMs in TxG and (2) the universal applicability of Matras' (1998) pragmatic detachability scale.

7 References

- Boas, H.C. 2003. Tracing Dialect Death: The Texas German Dialect Project. *Proceedings of the 28th Meeting of the Berkeley Linguistics Society*, ed. J. Larson and M. Paster, vol. 28, 387-398. Berkeley: Berkeley Linguistics Society.
- Boas, H.C. 2005. Texas German Dialect. *Germany and the Americas*, ed. T. Adam, 1029-1035. Santa Barbara: ABC-CLIO.
- Boas, H.C. 2006. *Best-practice Recommendations in Documentary Linguistics*. Manuscript, University of Texas at Austin.
- Brinton, L.J. 1990. The Development of Discourse Markers in English. *Historical Linguistics and Philology*, ed. J. Fisiak, 45-71. Berlin: de Gruyter.
- Brody, J. 1987. Particles Borrowed in Spanish as Discourse Markers in Mayan Languages. *Anthropological Linguistics* 29: 507-521.
- Clyne, M. 2003. *Dynamics of Language Contact*. Cambridge: Cambridge University Press.
- De Rooij, V.A. 2000. French Discourse Markers in Shaba Swahili Conversations. *The International Journal of Bilingualism*, 4: 447 – 467.
- Fuller, J. 2001. The Principle of Pragmatic Detachability in Borrowing: English-origin Discourse Markers in Pennsylvania German. *Linguistics* 39: 351-369.
- Fuller, J. 2003. The Influence of Speaker Roles on Discourse Marker Use. *Journal of Pragmatics* 35: 23-45.
- Gumperz, J. and E. Hernández-Chaves. 1971. Bilingualism, Bidialectalism, and Classroom Interaction. *Language in Social Groups: Essays by John J. Gumperz*, ed. A. Dil, 311-339. Stanford: Stanford University Press.
- Hall, J.N., and R. Schwartz. 1997. *Effective Perl Programming: Write Better Programs with Perl*. New York: Addison-Wesley.
- Hentschel, E. 1986. *Funktion und Geschichte deutscher Partikeln: Ja, Doch, Halt und Eben*. Tübingen: Max Niemeyer Verlag.
- Holmes, J. 1986. Functions of *You Know* in Women's Speech. *Language in Society* 15: 1-22.
- Jucker, A.H. 1993. The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics* 19: 435-452.
- Maschler, Y. 1994. Metalanguaging and discourse markers in bilingual conversation. *Language in Society* 23: 281-331.
- Maschler, Y. 1998. On the transition from code-switching to a mixed code. *Code-Switching in Conversation: Language, Interaction and Identity*, ed. P. Auer, 125-149. New York: Routledge.
- Matras, Y. 1998. Utterance Modifiers and Universals of Grammatical Borrowing. *Linguistics* 36: 281-331.
- Myers-Scotton, C. 2002. *Contact Linguistics*. Oxford: Oxford University Press.

- Östman, J.-O. 1981. *You Know: A Discourse Functional Approach*. Amsterdam: John Benjamins.
- Raith, J. 1992. Dialect mixing and/or code convergence: Pennsylvania German? *Diachronic Studies on the Languages of the Anabaptists*, ed. K. Burrige and W. Enninger, 152-165. Bochum: Brockmeyer.
- Salmons, J. 1990. Bilingual Discourse Marking: Code Switching, Borrowing, and Convergence in Some German-American Dialects. *Linguistics* 28: 453 – 480.
- Serra, C.O. 1998. Discourse connectives in bilingual conversation: the case of an emerging Italian-French mixed code. *Code-Switching in Conversation: Language, Interaction and Identity*, ed. P. Auer, 101-122. New York: Routledge.
- Schiffirin, D. 1985. Conversational Coherence: The Role of *Well*. *Language* 61: 640-667
- Schiffirin, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Schourup, L. 1985. *Common Discourse Particles in English Conversation*. New York: Garland.
- Schourup, L. 1999. Tutorial overview: discourse markers. *Lingua* 107: 227-265.
- Texas German Dialect Archive. 2005. [<http://www.tgdp.org>].
- Trudgill, P. 2004. *New-Dialect Formation*. Oxford: Oxford University Press.
- Weinreich, U. 1959. Is a Structural Dialectology Possible? *Word* 10, 388-400.
- Woolford, E. 1983. Bilingual Code-switching and Syntactic Theory. *Linguistic Inquiry* 14: 520-536.

Rule Replacement in Lingít (Tlingit): The Importance of Morphosyntax to Morphophonology¹

SETH CABLE

1 Introduction

This paper presents one case-study illustrating that attention to the morpho-syntactic structure of surface forms is sometimes critical to a proper understanding of seemingly morphophonological phenomena. Moreover, the unique advantages of the analysis developed here rest upon both the Distributed Morphology architecture and the existence of Fusion rules within that architecture. Thus, I argue that the phenomena described here provide

¹ I thank the following persons for their helpful discussion of earlier incarnations of this work: Karlos Arregi, Morris Halle, Heidi Harley, Brent Henderson, Alec Marantz, David Pesetsky, Norvin Richards, Donca Steriade and Jochen Trommer. Special thanks go out to Richard Dauenhauer, Roby Littlefield and all the members of the Tlingit Language and Culture Discussion List. This work has been much improved by its audiences at NELS36, TLS9, WSCLA10 and the MIT Phonology Circle. All errors are the fault of the author. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

The main argument of this paper may be summarized as follows. Within Lingít (Tlingit), the morphophonological exponents of perfective aspect and subject agreement are often ‘combined’ with surrounding prefixes into a single portmanteau morph. Under certain conditions, some portmanteau morphs in this system are ‘replaced’ by special forms. The environments affecting these portmanteau morphs do not constitute a phonological natural class, nor do the portmanteau morphs affected by these environments. However, both the environments and the morphs affected constitute morphosyntactic natural classes. If one attends to the morphosyntactic features combined by the morphs, and if one assumes that these features are combined via a *Fusion* rule, then a natural statement of the ‘portmanteau rule replacement’ becomes possible. The conceptual advantages of the analysis demonstrate that if one looks beyond the ‘phonological appearance’ of the alternations in question, and considers the morphosyntactic features being combined, then one can begin to see generalizations and connections that are otherwise hidden.

2 Verbal Portmanteau Allomorphy in Lingít

Lingít is a Na-Dene language spoken in Alaska and British Columbia.^{3, 4} Like its distant relatives in the Athabaskan family, verbal inflectional and derivational morphology is almost exclusively prefixal, and comparatively complex rules of contextual allomorphy can serve to drastically alter the underlying phonological form of the verbal prefixes. Certain of these alternations are given a constraint based analysis in Cable (2004).

Other alternations, however, appear to defy serious phonological analysis; this is especially the case with the aspectual and subject agreement morphemes. For certain combinations of subject agreement and aspect features, the phonological realizations of those features on the Lingít verb is a form that is not derivable from the usual underlying forms of the prefixes and the general phonology of the verbal prefix string. The targeted patterns of allomorphy may be described by the rewrite rule system in (1). The appendix to this paper collects a number of textually attested forms illustrating most of these alternations.

³ Throughout this paper, I restrict my discussion to the Northern dialect of Lingít. Its morphophonological differences from the Southern dialect are not of consequence to the analysis put forth here.

⁴ Pioneering grammatical study of Lingít can be found in, *inter alia*, Boas (1917), Naish (1966), Story (1966), Story & Naish (1973), Leer (1991), Dauenhauer & Dauenhauer (2000).

(1) Coalescence Phenomena (adapted from Story 1966, Leer 1991)

Morphemes Involved:

wu ‘perfective’
ee ‘second person singular subject’
ya ‘null series classifier, +I, -D’⁶
yi ‘second person plural subject’
Ci ‘(any) nonnull series classifier, +I’⁷

Rewrite Rules:⁵

a. *wu* + *ya* → *woo*
 b. *wu* + *ee* → *yi*
 c. *wu* + *ee* + *ya* → *yee*
 d. *yi* + *ya* → *yeey*
 e. *yi* + *Ci* → *yeeyCi*
 f. *wu* + *yi* → *yeey*
 g. *wu* + *yi* + *ya* → *yeey*

When presented with this set of facts, one’s natural inclination is to further simplify the system, perhaps by viewing some of the more complex alternations (1c, 1g) as the mere composition of some of the simpler ones. Before I present one proposal for doing just this, let us consider whether it is profitable to view the alternations above as lying within the phonological system of the language. After all, some of the alternations in (1) seem as if they might have a natural phonological basis (e.g. (1b)).

Despite the initial plausibility, however, it is probably best not to view these rules as forming a part of Lingít phonology. There are two facts which point to this conclusion. The first is that some of the coalescence alternations in (1) fail to apply under a number of *morphosyntactically* specified conditions. That is, there are certain conditions under which some alternations in (1) do not apply; these conditions do not form a phonologically natural class, though they do form a morphosyntactic natural class. Section 5 will discuss these facts in greater detail. The reader will see that a phonological construal of the alternations in (1) would have to assume they are sensitive to morphosyntactic properties of the word that would not normally be visible to the phonology.

The strongest reason to avoid a phonological analysis of these alternations, however, is that no such analysis would succeed in simplifying the statements in (1). For example, consider alternation (1c). Might it be possible to view (1c) as the composition of some of the other alternations in (1)? Under a phonological construal of the alternation, this is impossible. First, note that the sequence *yi* – *ya* surfaces as *yeey* (rule (1d)). Thus, (1c) cannot be the composition of (1d) and (1b). Next, note that the sequence *ee*

⁵ These rules are understood to apply when the morphs in question are directly adjacent.

⁶ For explanations of the terms ‘null series’, ‘nonnull series’ ‘+/-D’ and ‘+/-I’ in the context of Lingít verbal classifiers, see Leer (1991; section 4.1.1).

⁷ This description encompasses the following classifiers: *si*, *dzi*, *li*, *dli*, *shi*, *ji*

–*ya* otherwise surfaces as *iya*, via a regular shortening rule. Thus, alternation (1c) cannot be due to the application of some other rule in (1) to the phonological realization of *ee* – *ya*. The reader, of course, is invited to attempt more sophisticated phonological analyses of these alternations; the present author, however, has become frustrated in his attempts.

In this context, note also that four of the seven alternations under (1) have the form *yeey* as part of their output. Moreover, there seems to be no way to analyze the binary *yeey* alternations (1d), (1e) and (1f) as deriving from some more primitive phonological operations, and neither are these three binary alternations sufficient to derive the ternary *yeey* alternation (1g). Thus, a phonological analysis of the system in (1) would require four separate phonological rules, each of which produces the surface form *yeey*. A phonological analysis, therefore, provides no simpler statement of the distribution of the form *yeey* than the heterogeneous rule set in (1).

Although the system in (1) does not submit to phonological analysis, I will argue in the following sections that insight into its nature can be gained if we ignore its phonological appearance and pay closer attention to the morphosyntactic features being combined.⁸

3 Towards a Morphosyntactic Analysis

As a first step towards a more syntactic analysis of this system, let us note again that over half the ‘rules’ in (1) have the form *yeey* as part of their output. This is despite the wildly differing phonological and morphological properties of the assumed underlying forms. One might gather from this heterogeneous distribution that the form *yeey* is a morphological ‘default’, a morpheme which surfaces when the underlying morphosyntactic features cannot be mapped to any more specific phonological form (Bonet 1995, Halle & Marantz 1993). If we pursue this conception of the distribution of *yeey*, it will be possible to avoid a heterogeneous set of rules governing its appearance, such as appears in (1).

The system to be presented below is one that adopts this conception of *yeey*. It is divided into two components: a set of Fusion Rules governing the combination of morphosyntactic nodes, and a set of Vocabulary Rules governing the mapping between morphosyntactic nodes and phonological form. As a terminological point, then, the proposed system constitutes a ‘distributed morphology’ (Halle & Marantz 1993).

⁸ See Bonet (1995) for the pioneering study applying this approach to the analysis of Romance clitic clusters.

3.1 Fusion Rules

In this and the following section I make clear some of the architectural assumptions underlying the proposed formal analysis.⁹ The morphological system is conceived of as taking as input a structure composed purely of morphosyntactic features. Thus, the input to morphology is conceived of as a structure such as that represented in (2).

(2) [[[1st person] [plural]] [imperfect, 3rd]]

These structures are then manipulated and altered by various structure changing operations internal to the morphology. One of these operations is Fusion. Fusion combines two sets of morphosyntactic features (Halle & Marantz 1993). It is defined as in (3).

- (3) Fusion:
- (a) the result of Fusion to feature sets A B is the union of A and B
 - (b) Fusion can only apply to feature sets that are directly adjacent.

An illustrative morphological derivation employing Fusion is offered in (4). Note that Fusion of a node A containing features (a, b, c, d) and a node B containing features (e, f, g, h) produces a node C containing all eight features.

(4) INPUT [A a, b, c, d] [B e, f, g, h]
 FUSION of A and B [c a, b, c, d, e, f, g, h]

With this definition of Fusion in place, we can state the morphosyntactic rules in (5).

- (5) a. [Perf] and [AgrS, 2nd] are Fused
 b. [Perf] and [null, +I, -D] are Fused
 c. [AgrS, 2nd, pl] and [null, +I, -D] are Fused

The content of these rules will be later clarified via illustrative derivations. For now, let me state that these rules contain *minimal* descriptions of the nodes targeted by Fusion. For example, rule (5a) states that if *any* node containing the feature ‘Perf’ and *any* node containing the features ‘AgrS, 2nd’ are adjacent, then they are Fused.

⁹ For more details regarding ‘distributed morphologies’, see Halle & Marantz (1993), Harley & Noyer (1999). For reasons of space, I make here a number of simplifications (particularly regarding Fusion) that may distress savvy readers.

3.2 Vocabulary Rules

After Fusion and other structure changing operations have applied, the resulting structure is assigned a phonological form. This form is largely dependent upon the language's set of 'Vocabulary Rules'. These rules are instructions for phonologically realizing particular morphosyntactic feature combinations. A Vocabulary Rule can apply if its feature specification is *not inconsistent with* (i.e., is a subset of) the features on a given morphosyntactic node. For example, of the two Vocabulary Rules in (6), only the second can apply to the input morphosyntactic node.

- (6) INPUT [Feminine, 2nd, sg]
 Vocab Rule 1: [Masculine, 2nd, sg] ⇔ / gu / (can't apply)
 Vocab Rule 2: [Feminine, 2nd] ⇔ / ral / / ral /
 OUTPUT / ral /

In cases where more than one Vocabulary Rule may apply to a given node, a rule ordering – described presently – determines which applies.

The proffered set of Vocabulary Rules for Lingít are listed under (7).

- (7) a. [Perf, 2nd, sg, null, +I, -D] ⇔ / yee /
 b. [Perf, 2nd, sg] ⇔ / yi /
 c. [AgrS, 2nd, pl] ⇔ / yi / / ____[-I]
 d. [AgrS, 2nd, sg] ⇔ / ee /
 e. [AgrS] ⇔ / yeey /
 f. [Perf, null, +I, -D] ⇔ / woo /
 g. [Perf] ⇔ / wu /
 h. [null, +I, -D] ⇔ / ya /

The ordering of the rules in (7) is crucial for the correct operation of the system. It will usually be the case that a large number of Vocabulary Rules *could* apply to a given morphosyntactic node. In such cases, the ordering of the Vocabulary Rules determines which of the competing Vocabulary Rules *does* apply to the node. Of the Vocabulary Rules that may apply to it, a given morphosyntactic node undergoes that which appears highest in the ordering.

The reader will note that the ordering in (7) respects the standardly assumed 'Subset Principle' that a Vocabulary Rule be ordered after all those whose morphosyntactic specifications are supersets of its own. For more on the ordering of Vocabulary Rules, see the works cited in footnote 9.

- (10) [Perf] [AgrS, 2nd, pl] [null, +I, -D] FUSE, by (5a)
 [Perf, AgrS, 2nd, pl] [null, +I, -D] FUSE, by (5b) or (5c)
 [Perf, AgrS, 2nd, pl, null, +I, -D] SpellOut, by (7e)
 / yeey /
- [Perf] [AgrS, 2nd, pl] [null, +I, -D] FUSE, by (5c)
 [Perf] [AgrS, 2nd, pl, null, +I, -D] FUSE, by (5a) or (5b)
 [Perf, AgrS, 2nd, pl, null, +I, -D] SpellOut by (7e)
 / yeey /

I will talk the reader through one of the four derivations above; I assume that the graphical representation in (10) will render the other three clear enough. We assume that the input to morphology is the node containing *Perfective*, followed by the node containing *AgrS 2nd pl*, followed by the node containing *null, +I, -D*. To this input, Fusion Rule (5a) may apply, joining together the first two nodes in the sequence. Subsequently, Fusion Rule (5c) can now apply. The result is a single node containing all the features *Perf, AgrS 2nd pl, null +I -D*. To such a node, rule (7e) is the first Vocabulary Rule that can apply.

I will assume that it is now clear how derivations within this distributed morphology operate. Below I present the derivations the system provides for the other portmanteau allomorphs in (1), without accompanying prose.

- (11) Alternation (1b): *wu + ee* → *yi*
 [Perf] [AgrS, 2nd, sg] FUSE, by rule (5a)
 [Perf, AgrS, 2nd, sg] SpellOut, by rule (7b)
 / yi /
- (12) Alternation (1d): *yi + ya* → *yeey*
 [AgrS, 2nd, pl] [null, +I, -D] FUSE, by rule (5c)
 [AgrS, 2nd, pl, null, +I, -D] SpellOut, by rule (7e)¹¹
 / yeey /

¹¹ Rule (7c) cannot apply because its environmental condition is not met; the node is by necessity not adjacent to a -I classifier. The rule ordering in (7) entails that (7e) be used.

- (13) Alternation (1e): $yi + Ci \rightarrow yeeyCi$
 [AgrS, 2nd, pl] [s/l/sh, +I, -D] SpellOut, by rule (7e)¹²
 / yeey / [s/l/sh, +I, -D] SpellOut, by whatever¹³
 / yeey / / Ci /
- (14) Alternation (1f): $wu + yi \rightarrow yeey$
 [Perf] [AgrS, 2nd, pl] FUSE, by rule (5a)
 [Perf, AgrS, 2nd, pl] SpellOut, by rule (7e)¹⁴
 / yeey /

The derivations in (10), (12) – (14) illustrate the way in which the morpheme *yeey* behaves as a ‘default’ within this system. The morphosyntactic content of *yeey* is highly underspecified – all it signifies is the presence of subject agreement features. Its heterogeneous distribution is a direct result of its highly underspecified content, given the logic of the Subset Principle.

5 A Critical Complication

Thus far, we have seen that the system of Fusion Rules in (5) and Vocabulary Rules in (7) are sufficient to derive the pattern of allomorphy in (1). Besides being empirically adequate, this analysis captures the complex distribution of *yeey* by assigning it a single, highly underspecified environment. In addition, this morphosyntactic analysis receives further support from a curious set of conditions governing the portmanteau forms in (1).

Interestingly, alternations (1a) and (1c) do not apply under certain morphosyntactically specified conditions. Under these conditions, alternation (1a) is replaced with alternation (1a’) and alternation (1c) is replaced with alternation (1c’).

- (1a’) $wu + ya \rightarrow uwa$
 (1c’) $wu + ee + ya \rightarrow iya$

¹² No Fusion Rule can apply since the classifier is not of the null series. Moreover, rule (7c) again cannot apply because the node is not adjacent to a –I classifier.

¹³ Note that a left-to-right application of the Vocab Rules is required to derive forms with the 2nd pl prefix *yi*. Happily, there are no cases in which a right-to-left application of the Vocab Rules is essential to derive the correct output form. Given the evidence that the prefix string in a Na-Dene verb is leftward branching (Rice 2000), this provides additional evidence that Spell-Out proceeds in a bottom-up fashion (see Bobaljik 2000).

¹⁴ Rule (7c) cannot apply because the perfective prefix selects for a +I classifier. Hence, its environmental condition is never met.

The conditions under which (1a') and (1c') apply are the following.

(15) The Conditions Requiring Rule Alternates (1a') and (1c')¹⁵

- a. The verb is a member of the 'first conjugation'; its conjugation marker ('aspect prefix' in the terminology of Leer 1991) is the null prefix.¹⁶

Example:

Ách áyá a ká-t aa **wu-ya-át** → **uwa.át**¹⁷

so foc it top-to part. perf-cl-go

So they started over it. (Dauenhauer & Dauenhauer 1987; p. 68)¹⁸

- b. The perfective (*wu*) is directly preceded by an incorporated noun.

Example: (compare to example (26b) in Appendix)

...yoo haa **ka-wu-ya-nei** → **kaawanéi**¹⁹

part. us top-perf-cl-do

It happened to us. (D&D 1987; p. 82, line 3)

- c. The perfective is directly preceded by one of the following object agreement prefixes: second singular (*i*), second plural (*ye*), fourth (*ku*), third obviative (*a*).²⁰

Example: (compare to example (26b) in Appendix)

Wáa sáyá **i-wu-ya-nei** → **eewanei**²¹

how foc you-perf-ya-do

What happened to you? (D&D 1987; p. 124, line 28)

¹⁵ See Leer (1991; p. 177, 178, 185 - 202) and Story (1966; p. 115, 117).

¹⁶ Such verbs are referred to as 'Telic' by Leer 1991, the complement of this class being 'Atelic'. These verbs are referred to as 'K-Paradigmatic' by Story 1966, the complement of this class being 'L-Paradigmatic'. The term 'first conjugation' is introduced in Story & Naish (1973; p. 379).

¹⁷ In many places, the glosses I offer here are rather rough and oversimplified.

¹⁸ For reasons of space, I will henceforth use 'D&D' to abbreviate these authors' names.

¹⁹ A regular process of hiatus avoidance produces this surface form from the underlying form 'ka-uwa-nei'.

²⁰ Prior authors have referred to the *a* prefix as 'nonfocal'. Leer (1993) notes that the contrast between this prefix and the 'focal' 3rd Agr O (*ash*) seems quite similar to that between so-called 'obviative' and 'proximate' agreement in other languages. There are, however, other agreement prefixes that *a* contrasts with, which complicates its full analysis. See Leer (1993) for a rich discussion of the uses of these prefixes.

²¹ Hiatus avoidance produces this surface form from the underlying form 'i-uwa-nei'.

Although these facts may seem daunting at first, it will ultimately be shown that they provide crucial evidence supporting our morphosyntactic analysis. In brief, it will be argued that the morphosyntactic analysis provides more insight into the nature of these interesting irregularities than any phonological analysis.

Let us begin by asking two questions regarding the ‘rule replacement’ described in (15). The first question is ‘Why should all and only rules (1a) and (1c) be replaced in the environments specified? Why should these two rules behave as a class?’ The second question is ‘What commonalities unite the set of conditions in (15)? What, for example, do the prefixes in (15c) have in common such that all induce use of rules (1a’) and (1c’)?’ Ultimately, an account that can offer answers to these questions is more promising than one that cannot.

Consider now a phonological construal of the alternations in (1). What sort of answer could such an analysis provide to our first question? Unfortunately, it seems that no answer is forthcoming. After all, there don’t seem to be any *phonological* properties that unite alternations (1a) and (1c) to the exclusion of all the other alternations. Although both (1a) and (1c) involve the prefixes *wu* and *ya*, they aren’t the only alternations to do so; consider alternation (1g). It’s also rather unclear what phonological ‘operations’ the two alternations might exclusively share. Although it is not possible to prove a negative here, it remains rather difficult to see how alternations (1a) and (1c) can fall out as a natural class under a purely phonological construal of them.

Let us now ask what answer a phonological analysis of (1) might provide to our second question. Again, it seems that no answers are forthcoming. The prefixes in (15c) are not a phonologically natural class. The only phonological property that *i*, *yee*, *ku* and *a* have in common is that they are open syllables. However, there are other object agreement prefixes occupying the same ‘templatic position’ as those in (15c) that constitute open syllables, and which do *not* require the use of (1a’) and (1c’) – for example, the first person plural object agreement prefix *haa*. Once again, under a phonological analysis it remains unclear why the entities appealed to in (15) should group together.

It is thus doubtful that a phonological analysis of (1) could provide answers to our two questions regarding (15). On the other hand, some progress on these questions can be made if we adopt the morphosyntactic analysis developed in section 3. First of all, there immediately springs to mind a property uniting alternations (1a) and (1c): *these two alternations are the only ones to crucially rely upon Fusion Rule (5b)*. A quick inspection of the derivations in section 4 reveals that our morphosyntactic system

cannot derive either alternation (1a) or (1c) if appeal to rule (5b) is prevented, and that *no other alternation has this property*.

Let us, then, entertain the notion that the appearance of alternations (1a') and (1c') results from the inability for Fusion Rule (5b) to apply when the conditions in (15) obtain. Such an account would derive that the conditions in (15) would only affect the alternations in (1a) and (1c), and so would provide an interesting answer to our first question.

In the next section, we will develop in detail this nascent analysis of the 'rule replacement' in (15). This extension of our morphosyntactic analysis will provide answers to both our questions above, questions which seem to stymie any phonological analysis of the alternations. On these grounds, our full, final morphosyntactic analysis will be the superior account of the allomorphy in (1).

6 Morphosyntactic Analysis of the 'Rule Replacement'

Let us begin by amending the Fusion Rules so that they read as follows.

- (16) a. [Perf] and [AgrS, 2nd] are Fused
 b. [Perf] and [null,+I, -D] are Fused *unless* [Perf] is adjacent to:
 (i) [N]
 c. [AgrS, 2nd, pl] and [null, +I, -D] are Fused

Under the assumption that the appearance of alternations (1a') and (1c') is ultimately due to the 'suspension' of the Fusion operation in (5b), the 'unless' condition in (16b) directly builds into our system the sensitivity of the rule replacement to condition (15b). This will be seen in greater detail later on.

Consider now condition (15c). We have seen that a phonological analysis cannot characterize as a natural class the prefixes appealed to in condition (15c). Interestingly, when we turn our attention to the morphosyntactic features of these prefixes, we find that they *are* a natural class with respect to those features. Note that the prefixes listed in (15c) are all the object agreement prefixes *except* for 1st person singular *xat*, 1st person plural *haa*, and proximate third *ash*. Why should first person prefixes group together with proximate thirds? Let us adopt the notion that first person prefixes in Lingít are inherently 'proximate'. It is well known that, in languages witnessing obviate/proximate distinctions, local subjects tend to group together with proximates. Algonquian, for example, has many well known instances of this. Indeed, Halle & Marantz (1993) explore the possibility that local persons are obligatorily marked as 'proxi-

mate' in Potawatomi. Regarding languages within the Na-Dene family, Rice (2000; p. 220) notes that local persons in Athabaskan languages appear to be 'inherently topical.' Adopting the position that first persons in Lingít are inherently topical/proximate, we might restate the condition in (15c) to 'the perfective is directly preceded by a *non*proximate object agreement prefix.'^{22, 23} We might incorporate this version of condition (15c) into our Fusion Rules by making the following amendment.

- (17) a. [Perf] and [AgrS, 2nd] are Fused
 b. [Perf] and [null,+I, -D] are Fused *unless* [Perf] is adjacent to:
 (i) [N] ; (ii) [- Topic]
 c. [AgrS, 2nd, pl] and [null, +I, -D] are Fused

At this point, however, the reader might note an interesting redundancy within condition (17b). Recall that one of the effects of noun incorporation on discourse structure is the *detopicalization* and *backgrounding* of the information contributed by the noun (Mithun 1999; p. 46). Incorporated Ns are thus by necessity nontopical, and statement (17bii) can be made to cover (17bi).

- (18) a. [Perf] and [AgrS, 2nd] are Fused
 b. [Perf] and [null,+I, -D] are Fused *unless* [Perf] is adjacent to:
 (i) [- Topic]
 c. [AgrS, 2nd, pl] and [null, +I, -D] are Fused

We find, then, that our morphosyntactic analysis of (1) – unlike a phonological analysis – can provide an interesting answer to the second of our questions in section 4. Conditions (15b) and (15c) ultimately reduce to a

²² Of course, this proposal begs the question why *second* person is not treated by Lingít as inherently proximate. One answer might be that this fact simply reflects the tendency for the 'animacy hierarchy' to differ slightly across languages (Comrie 1989; chapter 9). Under this view, the Lingít system represents a more extreme version of the systems found in Algonquian and Athabaskan, in which only the most highly ranked local person – first – qualifies as proximate.

²³ According to the analyses of Story (1966) and Leer (1991), there are other prefixes occupying the same 'templatic position' as the seven aforementioned object agreement markers: the reflexive prefix *sh*, indefinite object *at* and partitive object *aa*. As (15) indicates, these prefixes do not require use of rules (1a') and (1c'). Plausibly, these three prefixes differ from object agreement prefixes by their lack of referentiality. This lack of referentiality could entail that they do not bear *any* topic features – either positive or negative – and so would they would be expected under (17b) not to require use of alternations (1a') and (1c'). On the other hand, it is also possible that these prefixes do not, in fact, occupy the same morphosyntactic position as the object agreement markers, and so do not count as 'adjacent' for the purposes of rule (17b).

single condition requiring (1a') and (1c') when [Perf] follows material marked as [- Topic].

Finally, let us turn our attention to condition (15a). Our guiding 'intuition' requires that the conjugational class of a verb be able, somehow, to interrupt the Fusion of [Perf] and [null +I -D]. Some insight into the exact mechanics of this interaction can be gained by considering the 'morphological template' of the Lingit verb.

(19) Portion of the Lingit Verbal Prefix Template (based on Leer 1991)

...AgrO-IncorpN-Conjugation1-Irrealis-Conugation2-Aspect-Dist-...
 { ga } { u } { na, ga, Ø } { wu }

Following the proposals in Leer (1991), the null, first conjugation prefix 'Ø' always appears in a position directly preceding the position of the perfective prefix.²⁴ Therefore, we might assume that in first conjugation verbs, the node containing the feature [Perf] is always adjacent to a node containing the feature [Conj 1st].²⁵ We might then incorporate condition (15a) into our formal system by amending Fusion Rule (18b) in the following way.

- (19) a. [Perf] and [AgrS, 2nd] are Fused
 b. [Perf] and [null,+I, -D] are Fused *unless* [Perf] is adjacent to:
 (i) [- Topic] ; (ii) [Conj 1st]
 c. [AgrS, 2nd, pl] and [null, +I, -D] are Fused

The 'unless' condition in rule (19b) has as its consequence that [Perf] and [null, +I, -D] *cannot* be fused when any of the conditions in (15) obtain. The result is that alternations (1a) and (1c) – and only those alternations – will fail to occur if and only if those conditions apply. Thus, our morphosyntactic system is quite close to deriving the 'rule replacement'

²⁴ The motivation for placing the null prefix in this position is not strong. Leer (1991) notes that the three prefixes occupying our 'Conjugation2' are in complementary distribution. This argument, however, is weakened by the fact that each of the Conjugation2 prefixes is also in complementary distribution with the prefix *ga*, in the Conjugation1 position. Neither Story (1966) nor Story & Naish (1973) recognize a null conjugation prefix, only a conjugation class that is not signaled by a formal prefix.

²⁵ A potential complication arises from the fact that no verb of *any* conjugational class appears containing an overt conjugational prefix in its perfective mode. However, this might simply be due to a readjustment rule that requires all conjugational prefixes to surface as null in the environment of the feature [Perf].

introduced in Section 5. As it is presently structured, however, our system does not produce the correct outputs when the conditions in (15) obtain. The following derivations illustrate.

(20) Alt. (1a’): $\{-\text{Topic}, [\text{Conj } 1^{\text{st}}]\}$ -wu-ya \rightarrow $\{-\text{Topic}, [\text{Conj } 1^{\text{st}}]\}$ -uwa

$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf] [null,+I,-D]	SpellOut by rule (7g)
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / [null,+I,-D]	SpellOut by rule (7h)
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / / ya /	Regular Phonology
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / / wa /	

(21) Alt. (1c’): $\{-\text{Topic}, [\text{Conj } 1^{\text{st}}]\}$ -wu-ee-ya \rightarrow $\{-\text{Topic}, [\text{Conj } 1^{\text{st}}]\}$ -iya

$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf] [AgrS 2 nd sg][null,+I,-D]	FUSE, by (19a)
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf AgrS 2 nd sg][null +I -D]	SpellOut, by (7b)
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / yi / [null +I -D]	SpellOut, by (7h)
$\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / yi / / ya /	

These derivations demonstrate that our system presently outputs *wuwa* instead of the correct output *uwa*, and *yiywa* instead of the correct output *iya*. Note, however, that these incorrect outputs are tantalizingly close to the correct outputs. Indeed, all that is needed to convert them into the correct outputs is the rather simple rule of glide deletion in (22).

(22) Lingít Successive Glide Onset Deletion:

$$[C_{\text{glide}} V]_{\sigma} [C'_{\text{glide}} V]_{\sigma} \rightarrow [V]_{\sigma} [C'_{\text{glide}} V]_{\sigma} ; \text{ where } C = C'$$

This rule – possibly rooted in the OCP – deletes glides when they occupy the onset of a syllable followed by a syllable with an identical glide in its onset. If this rule is appended to the end of the derivations above, our system derives the correct outputs.

(23) $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf] [null,+I,-D] SpellOut by rule (7g)
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / [null,+I,-D] SpellOut by rule (7h)
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / / ya / Regular Phonology
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / wu / / wa / Glide Deletion
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / u / / wa /

(24) $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf][AgrS 2nd sg][null,+I,-D] FUSE, by (19a)
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ [Perf AgrS 2nd sg][null +I -D] SpellOut, by (7b)
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / yi / [null,+I,-D] SpellOut, by (7h)
 $\{-\text{Topic }], [\text{Conj } 1^{\text{st}}] \}$ / yi / / ya / Glide Deletion

{ [-Topic], [Conj 1st] } / i / / ya /

Thus, we find that the amended system of Fusion Rules in (19), the system of Vocabulary Rules in (7) and the single rule of glide deletion in (22) is sufficient to derive the portmanteau allomorphy described in (1), as well as the specially conditioned, ‘alternative’ allomorphy described in section 5.²⁶

7 Conclusion

Beyond its demonstrated empirical adequacy, the morphosyntactic analysis we have developed here has several conceptual advantages over a purely phonological analysis of the alternations in (1). First, it is able to provide an elegant, unitary account of the heterogeneous distribution of *yeey* via postulation of a single, highly underspecified Vocabulary Rule. Secondly, it is able to characterize alternations (1a) and (1c) as a ‘natural class’, and can derive their exclusive sensitivity to the conditions specified in (15). Finally, it is able to capture many of the conditions in (15) under a single generalization, one that appeals to the morphosyntactic features of the prefixes in question.

The concept of a ‘distributed morphology’ thus receives interesting support from the system of Lingít portmanteau allomorphy in (1). It is only within a ‘late insertion’ model of morphology employing Fusion rules that this system can be further analyzed and the results listed above be obtained. Note that in a ‘Minimalist’ distributed morphology like that developed in Trommer (1999, 2003), the absence of Fusion rules would prevent one from grouping alternations (1a) and (1c) as a natural class.²⁷ Rather, each alternation would have to be due to its own distinct rules of zero insertion and contextual allomorphy. Similarly, a structureless approach to morphology such as the ‘A-Morphous Morphology’ of Anderson (1992) could not easily group (1a) and (1c). The morphological input to an A-Morphous Morphology is an already ‘fully fused’ unit, and there is never a level at which morphosyntactic features are grouped in separate structural units from one another. Thus, there is no way within that framework to replicate the success-

²⁶ There is no independent evidence for this rule of glide deletion in Lingít. Nevertheless, Cable (2004) proposes the existence of several prosodic domains mapped to the Lingít verbal prefix string. It is, at least, reassuring to note that the rule in (22) may be consistently added to the phonology of one of those domains (the ‘Inner Prefix Domain’).

²⁷ Within Trommer’s system, apparent cases of Fusion are all analyzed as involving the insertion of a special null morpheme at one position, followed by concomitant rules of contextual allomorphy altering the forms of adjacent prefixes. Thus, all putative cases of Fusion are given an analysis akin to that provided in Halle & Marantz (1993) for English ablaut past tense forms such as ‘gave’.

ful notion that (1a) and (1c) form a class with respect to the structural manipulations required to group their underlying syntactic features.²⁸

The portmanteau allomorphy of Lingít provides a particularly striking object lesson in the relevance of morphosyntax to morphophonology. Here, as in so many other cases, if one looks beyond the phonological ‘appearance’ of the alternations, and considers the morphosyntactic features being combined, one can begin to discover generalizations and connections that are otherwise hidden.

8 Appendix

These textual examples witness most of the alternations discussed above.

(25) Regular surfacing of *wu*

Jilkáat aa-x̣ has **wu-si-tee** → **wusitee**
 Chilkat it-of they perf-cl-be
They became Chilkats. (D&D 1987; p. 68, line 117)

(26) Alternation (1a): *wu + ya* → *woo*

a. s'eenáa yaakw ax̣ jee **yéi wu-ya-tee** → **yéi wootee**
 seine boat my hand so perf-cl-be
I had a seine boat. (D&D 1987; p. 72, line 17)

²⁸ The portmanteau system in (1) also raises another interesting challenge to an ‘A-Morphous Morphology.’ Recall the description that alternation (1a) only applies when the forms ‘wu’ and ‘ya’ are directly adjacent to one another (footnote 5). When prefixes intervene between these forms, the portmanteau morph does not arise. Thus, the ‘underlying form’ *wu-too-ya-aat* ‘we left’ surfaces as *wu-tu-wa-aat* (via a regular assimilation rule), not *woo-tu-aat* or *tu-woo-aat*. Within an ‘A-Morphous Morphology’, the lack of morphosyntactic structure entails that one cannot state this ‘position based’ generalization regarding when *wu* and *ya* surface faithfully together. Rather, one must state feature based generalizations, such as ‘portmanteau morph (1a) does not appear when subject is 1st person’. However, the prefixes that linearly intervene between ‘wu’ and ‘ya’ are morphosyntactically heterogeneous. For example, the distributive prefix ‘daga’ also intervenes between ‘wu’ and ‘ya’. When this occurs, the output is again *wu-daga-a* (via a regular elision rule), not *woo-daga* or *daga-woo*. An A-Morphous Morphology must therefore state a disjunctive condition, like ‘portmanteau morph (1a) does not appear when subject is 1st person or action distributes over objects.’ Not only are such disjunctive conditions to be dispreferred, they also miss the obvious fact that the features which force the faithful appearance of *wu* and *ya* together are *exactly* those which are mapped to phonological forms linearly positioned between them.

Note that the proffered Distributed Morphology analysis captures the ‘position based’ generalization by means of the condition that nodes can only Fuse when they are directly, linearly adjacent (see (3b)).

b. kaawayí-x' **yóo wu-ya-nei** → **yóo woonei**
 above-to part. perf-cl-do
They went into the air. (D&D 1987; p. 78, line 122)

(27) Alternation (1b): *wu + ee* → *yi*

wu-ee-si-kóo wéit'át kookénaa → **ysisikóo**
 perf-2ndsg-cl-know thing messenger
You know what a messenger is. (D&D 1987; p. 232, line 286)

(28) Alternation (1c): *wu + ee + ya* → *yee*

sakwnéin **wu-ee-ya-xoox** → **yee~~xoox~~**
 bread perf-2ndsg-cl-ask
You asked them for bread. (D&D 1990; p. 186, line 15)

(29) Alternation (1d): *yi + ya* → *yey*

yeedát áwé **yi-ya-téen** → **yeytéen**
 now foc 2ndpl-cl-can.see
Now you can see. (D&D 1990; p. 176, line 18)

(30) Alternation (1e): *yi + Ci* → *yeyCi*

wooch **yi-dzi-xán** → **yeydzixán**
 recip. 2ndpl-cl-love
You care for each other. (D&D 1987; p. 104, line 456)

(31) Alternation (1f): *wu + yi* → *yey*

wu-yi-si-kóo yee kaani yán → **yeysikóo**
 perf-2ndpl-cl-know your brothers-in-law
You all know your brothers in law. (D&D 1990; p. 238, line 66)

(32) Alternation (1g): *wu + yi + ya* → *yey*

du jee-t **wu-yi-ya-tée** → **yeytée**
 his hand-to perf-2ndpl-cl-be
You gave it to him. (D&D 1990; p. 176, line 10)

(33) Alternation (1a'): $wu + ya \rightarrow uwa$

(see section 4)

(34) Alternation (1c'): $wu + ee + ya \rightarrow iya$

neil **wu-ee-ya-tée** \rightarrow **iyatée**
 house perf-2nd-cl-throw
You threw it in the house. (D&D 1987; p. 222, line 86)

References

- Anderson, S. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- Boas, F. 1917. *Grammatical Notes on the Language of the Tlingit Indians*. University of Pennsylvania University Museum Anthropological Publications 8:1.
- Bobaljik, J. 2000. The Ins and Outs of Contextual Allomorphy. *University of Maryland Working Papers in Linguistics 10*, ed. K.K. Grohmann and C. Struijke. Maryland.
- Bonet, E. 1995. Feature Structure of Romance Clitics. *Natural Language and Linguistic Theory* 13: 607–47.
- Cable, S. 2006. A Syncope in the Verbal Prefixes of Tlingit: Meter and Surface Phonotactics. *Lincom Studies in Native American Linguistics*. Muenchen: Lincom Europa.
- Comrie, B. 1989. *Language Universals & Linguistic Typology*. Chicago: The University of Chicago Press.
- Dauenhauer, N.M. & R. Dauenhauer. 1987. *Classics of Tlingit Oral Literature Volume 1: Haa Shuka, Our Ancestors: Tlingit Oral Narratives*. Seattle: University of Washington Press.
- Dauenhauer, N.M. & R. Dauenhauer. 1990. *Classics of Tlingit Oral Literature Volume 2: Haa Tuwunáagu Yis, For Our Healing Spirit: Tlingit Oratory*. Seattle: University of Washington Press.
- Dauenhauer, N.M. & R. Dauenhauer. 2000. *Beginning Tlingit*. Juneau, Alaska: Sealaska Heritage Foundation Press.
- Halle, M. & A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. *The View from Building 20*, ed. K. Hale and S. J. Keyser. Cambridge, MA: MIT Press.
- Harley, H. & R. Noyer. 1999. State of the Article: Distributed Morphology. *GLoT* 4.4: 3–9.

- Leer, J. 1991. *The Schetic Categories of the Tlingit Verb*. Doctoral dissertation, University of Chicago.
- Leer, J. 1993. The Tlingit Anaphoric System and its Roots in Na-Dene. MS. Alaska Native Language Center.
- Mithun, M. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.
- Naish, C. 1966. *A Syntactic Study of Tlingit*. Language Data: Amerindian Series 6. Dallas, Texas. Summer Institute of Linguistics.
- Rice, K. 2000. *Morpheme Order and Semantics Scope: Word Formation in the Athapaskan Verb*. Cambridge: Cambridge University Press.
- Story, G. 1966. *A Morphological Study of Tlingit*. Language Data: Amerindian Series 7. Dallas, Texas: Summer Institute of Linguistics.
- Story, G. & C. Naish. 1973. *Tlingit Verb Dictionary*. Alaska Native Language Center. University of Alaska.
- Trommer, J. 1999. Morphology Consuming Syntax's Resources: Generation and Parsing in a Minimalist Version of Distributed Morphology. *Proceedings of the ESSLI Workshop on Resource Logics and Minimalist Grammars*. Utrecht.
- Trommer, J. 2003. Feature (Non-)Insertion in a Minimalist Approach to Spellout. *Proceedings of CLS 39: Main Session*. Chicago.

Number Marking in Maltese Nouns

KARINE DAVID

1 Introduction

Maltese is a Semitic language. It is one of the two official languages of the Maltese archipelago (along with English). Its particularity resides in a morphology, a syntax, and a morphosyntax partially inherited from Arabic. The lexicon is composed for 70 % of Maghrebin Arabic, 20 % of Italian, and 10 % of loan words particularly English. An other particularity is that Roman characters are used to write Maltese.

In this work we concentrate on noun morphology and especially on the realization of number in Maltese nouns. Aquilina (1965), Borg & Azzopardi-Alexander (1997), Fenech (1996), and Sutcliffe (1936) distinguish three realizations of number for the nouns in Maltese: singular, dual, and plural. The singular form is used to express one entity, the dual form is used to express two entities, and the plural form is used to express three entities and more. So the three numbers: one entity, two entities, and three entities and more can be expressed by three different morphological forms: singular, dual, and plural.

But actually not every noun possesses the three forms: singular, dual, and plural. Most of them only have two, and there is no connection between the three forms and the three semantic numbers; the singular form is not always used to express one entity, the dual form is not always used to ex-

press two entities, and the plural form is not always used to express three entities and more.

First we present the data and show different configurations between the three forms. Then we analyze different hypotheses to propose finally our own analysis to explain the realization of the three numbers in Maltese nouns.

2 The data

2.1 The classification

In Maltese we can distinguish three classes of nouns.

The nouns of the Class 1 & 2 use only two of the three inflected forms globally available. And the Class 3 uses the three inflected forms.

Class 1: nouns with two forms: singular - plural

Nouns of the Class 1 possess two forms: a singular form to express one entity and a plural form to express more than one entity (two entities or more than two entities).

	One entity	Two entities	Three entities and more
CLASS 1	<i>saltna</i> kingdom	<i>saltⁿⁱet</i> kingdom-PL	<i>saltⁿⁱet</i> kingdom-PL

Table 1: Class 1

For these nouns the plural form is used to refer to more than one entity as in French or in English.

Class 2: nouns with two forms: singular - dual

Nouns of the Class 2 possess two forms: a singular form and a dual form to express a plural number.

	One entity	Two entities	Three entities and more
CLASS 2	<i>id</i> hand	<i>id^{ej}n</i> hand-DU	<i>id^{ej}n</i> hand- DU

Table 2: Class 2

Authors underline that nouns of this class are names of body parts which are pairs, and that suffixed forms are not considered by Maltese

speakers as forms expressing precisely two entities, but as forms expressing a plural number (more than one entity). Fenech (1996) characterizes the suffix of dual in this case as a ‘plural allomorph’ and gives an example of a body part which is not a pair to illustrate clearly how the dual form can express a plural number.

1. a. *difer*
 nail (of finger or toe)
- b. *difr -ejn*
 nail -DUAL
 nails

Class 3: nouns of three forms: singular -dual - plural

Nouns of the Class 3 possess three forms: a singular form, a dual form to express precisely a number of two entities, and a plural form to express a number of three entities and more.

	One entity	Two entities	Three entities and more
CLASS 3	<i>jum</i> day	<i>jumejn</i> day-DU	<i>ijiem</i> day-PL

Table 3: Class 3

Nouns possessing three numbers are semantically related to time, number, old Maltese weights, old Maltese measures, old Maltese coins, basic food, and some familiar objects.

In this class the dual form expresses precisely two entities. Fenech (1996) calls it ‘real dual’.

We observe that the dual form of Class 3 is similar to the dual form of Class 2, and that the plural form of Class 3 is similar to the plural form of Class 1. So nouns in Class 3 possess a paradigm with three forms which are strikingly similar to the non-singular forms of Class 1 and 2.

	ONE ENTITY	TWO ENTITIES	THREE ENTITIES AND MORE
Class 1	Singular	Plural	Plural
Class 2	Singular	Dual	Dual
Class 3	Singular	Dual	Plural

Table 4: Recapitulation of the forms

2.2 The realization of the marks

In this section we analyze the marking of the singular, the dual, and the plural. Singular is unmarked, dual is marked by suffixation, and plural is marked by suffixation or by other processes.

Singular

The singular form of the noun is the unmarked form which is employed as the basis to form the dual and the plural.

Dual

The dual form is formed by addition of the suffix [-*ejn*] or [-*ajn*] to the base or to the singular form. [-*ejn*] or [-*ajn*] are two forms of the same suffix. The variation between the linking vowel [-*e-*] and [-*a-*] is phonologically conditioned:

Rule: $e \rightarrow a / C_{[+back]} \text{ ____}$

2. *saq -ajn*
[saʔ-ajn]
foot-DUAL
3. *riǵl - ejn*
[rid3l- ejn]
leg -DUAL

We can find two variants for the dual suffix: [-*ej-*] and [-*aj-*], which are the reflex of the full forms [-*ejn*] and [-*ajn*] after the functioning of an apocope because of a further suffix, namely the attached pronominal pronoun. It is impossible to find a full form of the dual suffix preceding the attached pronominal pronoun.

- | | | |
|----|---|--|
| 4. | a. <i>saq- <u>aj</u> - <u>h</u></i>
foot-DUAL-POSS
his feet | b.* <i>saq - <u>ajn</u> - <u>h</u></i>
foot-DUAL-POSS |
| 5. | a. <i>riǵl- <u>ej</u> - <u>h</u></i>
leg-DUAL-POSS
his legs | b.* <i>riǵl- <u>ejn</u> - <u>h</u></i>
leg-DUAL-POSS |

- | | | |
|----|----------------------|--|
| 6. | a. <i>id</i>
hand | b. <i>id -<u>ejn</u></i>
hand-DUAL
hands |
|----|----------------------|--|

Plural

According to Sutcliffe, Aquilina, and Fenech, the plural of Maltese nouns can be classified into classes: suffixed plurals, broken plurals, and irregular plurals.

We find different types of suffixes to form the plural. Suffixes from Semitic origin:

- | | | |
|-----|---------------------------------|---|
| 7. | a. <i>qassis</i>
priest | b. <i>qassis<u>in</u></i>
priest-PL
priests |
| 8. | a. <i>saltna</i>
kingdom | b. <i>saltn<u>iet</u></i>
kingdom-PL
kingdoms |
| 9. | a. <i>sid</i>
owner | b. <i>sidi<u>en</u></i>
owner-PL
owners |
| 10. | a. <i>ħaddied</i>
blacksmith | b. <i>ħaddied<u>a</u></i>
blacksmith-PL
blacksmiths |

Suffixes from Italian:

- | | | |
|-----|-------------------------------|--|
| 11. | a. <i>ġurnal</i>
newspaper | b. <i>ġurnal<u>i</u></i>
newspaper-PL
newspapers |
|-----|-------------------------------|--|

Suffixes from English:

- | | | |
|-----|---------------------------------|---|
| 12. | a. <i>kompjuter</i>
computer | b. <i>kompjuter<u>g</u></i>
computer-PL
computers |
|-----|---------------------------------|---|

Broken plurals mean that they involve an internal change:

- | | |
|-----------------------------|---|
| 13. a. <i>but</i>
pocket | b. <i>bwiet</i>
pocket-PL
pockets |
|-----------------------------|---|

Irregular plurals or suppletive plurals involve a different radical from that of the singular:

- | | |
|-----------------------------|-------------------------------------|
| 14. a. <i>mara</i>
woman | b. <i>nisa</i>
woman-PL
women |
|-----------------------------|-------------------------------------|

2.3 Summary

There are different ways to form the dual and the plural, but in all the cases we meet a problem with the configuration: Singular-Plural-Plural of the Class 1 which uses the plural forms to express entities above 1 and with the configuration: Singular-Dual-Dual of the Class 2 which uses the dual form to express two entities, but expresses three entities and more too.

3 Analysis

First we explore hypotheses to find an analysis making clear the situation of the three numbers realization on the Maltese nouns, considering the occurrence of the three numbers and considering that very few nouns possess one particular form for each number. And finally we propose a solution.

3.1 Exploring hypotheses

H 1: One could postulate a three slot paradigm for nouns of the three classes with two defective classes (1&2).

	Singular	Dual	Plural
Class 1	+		+
Class 2	+	+	
Class 3	+	+	+

Table 5: Paradigm of three slots

It would mean that nouns of Class 1 could not express two entities and nouns of Class 2 could not express more than two entities. This is clearly not the case:

15. *it-tnejn iffal*¹
 det-two children
 the both children

16. *l-ghaxart idejn*
 det- ten hands
 the ten hands

In 15. a nominal phrase for a noun of Class 1 expresses two entities and in 16. a nominal phrase of a noun of Class 2 expresses ten entities.

H 2: ‘Number for Maltese nouns is just an opposition of singular versus non-singular’ as proposed by Corbett (2000). According to Corbett the noun paradigm has two slots. Class 3 is a deviant category, another type of noun, which possesses a minor number. On the one hand, this makes sense of the majority of nouns and clearly defines Class 3 as an exception. On the other hand, it does not account for the identity between the non-singular forms of Class 1 and 2 with the dual and plural forms of Class 3.

	Singular	Non-singular
Class 1	<i>saltna</i>	<i>saltniet</i>
Class 2	<i>id</i>	<i>idejn</i>

Table 6: Singular/non-singular

	Singular	Dual	Plural
Class 3	<i>jum</i>	<i>jumejn</i>	<i>ijjem</i>

Table 7: Class 3

H 3: As it is uncommon to find several paradigms for the members of the same grammatical category, especially with identical forms, we postulate that each noun of each class possesses three forms but that the forms are not always different. Note that the similarity be-

¹ The form of the plural is normally *tfal* but the euphonic *i* is present because a word can not begin with two consonants if the final letter of the previous word is a consonant.

	SINGULAR	DUAL	PLURAL
Class 1	<i>saltna</i>	<i>salt<u>ni</u>et</i>	<i>salt<u>ni</u>et</i>
Class 2	<i>id</i>	<i>id<u>e</u>jn</i>	<i>id<u>e</u>jn</i>
Class 3	<i>jum</i>	<i>jum<u>e</u>jn</i>	<i>ij<u>i</u>em</i>

Table 9: Solution

In the framework of Stump (2001) we could modelize this proposal with a single block, called block 1, containing the following five rules:

Block 1:

- $X_{N,\sigma}: \{\text{NB Sg}\} \rightarrow X$
- $X_{N,\sigma}: \{\text{NB Du}\} \rightarrow X \oplus \text{ejn}$
- $X_{N,\sigma}: \{\text{NB Pl}\} \rightarrow X \oplus \text{iet}$
- $X_{N\text{-CL1},\sigma}: \{\text{NB Du}\} \rightarrow \langle X, \sigma / \{\text{NB Pl}\} \rangle: 1$
- $X_{N\text{-CL2},\sigma}: \{\text{NB Pl}\} \rightarrow \langle X, \sigma / \{\text{NB Du}\} \rangle: 1$

The first rule specifies that the singular of a noun is the basic form (unmarked). The second rule forms the dual of a noun in suffixing [-ejn] to the base. The third rule forms the plural in suffixing [-in] to the base. And the two last rules are rules of referral which correspond respectively to:

- to form the dual of the nouns of Class 1 refer to the realization of the plural number in block 1
- to form the plural of the nouns of Class 2 refer to the realization of dual number in block 1.

The analysis explains the distribution of the three numbers but does not account for irregular plurals. This could be accommodated by using stem-space developed by Bonami & Boyé (2003) in the following of Pirelli and Battista (2000). Each noun would possess two stems: Stem1 to form the singular and the dual and Stem2 to form the plural.

SINGULAR	DUAL	PLURAL
Stem1	Stem1 \oplus ejn	Stem2

Table 10: Forms of the stem-space of the nouns in Maltese

4 Conclusion

The proposal presented in this paper gives an identical paradigm for every noun if we use the rules of referral. Moreover, Fenech (1996) and Corbett (2000) would agree on this analysis, and it makes sense of the idea that a class would possess a homogenous paradigm. However it would be more convenient if Class 1 would be unmarked and if Class 3 would be marked.

5 References

- Aquilina, J. 1965. *Teach Yourself Maltese*. London : First Edition : Hodder Headline Plc.
- Blanc, H. 1970 a. Dual and pseudo-dual in the Arabic dialects, *Language* 46 : 42-57.
- Bonami, O., Boyé, G. 2003. Supplétion et classes flexionnelles dans la conjugaison du français, *Langages* 152 : 102-126.
- Borg, A., Azzopardi-Alexander, M. 1997. *Maltese, Descriptive Grammars*. Series Editor : Bernard Comrie : Université du Sud de la Californie.
- Corbett, G. G. 2000. *Number*. Cambridge : Cambridge University Press.
- Fenech, E. 1996. Functions of the Dual Suffix in Maltese, *Rivista di Linguistica* 8.1. : 89-99.
- Hetzron, R. 1997. *The Semitic Languages*. New York : Routledge.
- Hockett, C. F. 1958. *A Course on Modern Linguistics*. New York : The McMillan Compagny.
- Jakobson, R., Fant, G. C. and Halle, M. 1952. *Preliminaries to Speech Analysis. The distinctive features and their correlates*. Edition 1967, Cambridge : The MIT Press.
- Lipinski, E. 1997. *Semitic Languages Outline of a Comparative Grammar*. Louvain.
- Martinet, A. 1965 a. *Eléments de linguistique générale*. Paris : Armand Colin.
- Matthews, P. H. 1974. *Morphology*. 2^d ed., 1991, Cambridge : Cambridge University Press.
- Pirrelli, V., Battista, M. 2000. The paradigmatic dimension of stem allomorphy in italian verb inflection, *Rivista di linguistica* 2 (12) : 307-380.
- Plank, F. 1996. Domains of the dual, in Maltese and in general, *Rivista di Linguistica* 8.1. : 23-40.
- Stump, G. 2001. *Inflectional Morphology*. Cambridge : Cambridge University Press.
- Sutcliffe E. D., 1936, *A Grammar of the Maltese Language*, Valletta, Progress Press.
- Zwicky A. M., 1991, Paradigms - The Economy of Inflection, *Empirical Approaches to Language Typology* 9 : 113-131.

A Domain-Based Approach to 2P Clitics in Pashto

ASCANDER DOST

1. Overview

Second position (2P) clitic phenomena have provided contemporary linguistics with a rich testing ground for the autonomy of grammatical components, as their placement and ordering often seems to require reference to both syntax and prosody. 2P clitics in Pashto frame these questions in a particularly interesting way; the constraints governing their placement cut across both syntax and prosody, and *at both* the lexical and sublexical levels. In this paper, I discuss a recent analysis of Pashto 2P clitics in Roberts (2000) and address some issues facing the proposal. I then present an alternative analysis within the formalism of Head-Driven Phrase Structure Grammar (Pollard and Sag, 1987, 1994, Sag and Wasow, 1999, Ginzburg and Sag, 2001, henceforth, HPSG), augmented with the word order domains of Kathol (1995).

2. Background

Pashto 2P clitics include pronominal, verbal and adverbial elements (see FIGURE 1) and display properties commonly associated with postlexical clitics; namely they are prosodically dependent on an adjacent prosodic element, exhibit a low degree of selection with respect to the syntactic category of their hosts, and take wide scope with respect to coordinated hosts (see Zwicky and Pullum (1983) for further discussion of the clitic/affix distinction). Tegey (1977) formulates the generalisation that 2P clitics appear after the first *stress-bearing* phrasal constituent

<i>Pronominal</i>		<i>Modal</i>	<i>Adverbial</i>
me	1SG	ba FUT, <i>can, may</i>	xo <i>indeed, really</i>
de	2SG	de <i>should, must</i>	no <i>then</i>
ye	3SG, 3PL		
am, mo	1PL, 2PL		

FIGURE 1 Inventory of Pashto 2P clitics

in the Pashto clause. As the examples in (1) illustrate, placement of 2P clitics may not result in discontinuous syntactic phrases (clitics here and throughout shown in boldface).

- (1) a. [aʃa xɑysta p y la]_{NP} **me** w lida
that pretty girl cl.1SG PF.see.PST.3SG.FEM
I saw that pretty girl.
- b. *aʃa **me** xɑysta p y la w lida
- c. [p  ʈop k]_{PP} **ye** zmaray wuw ft 
with gun cl.3SG lion PF.shoot.PST.3SG
He shot a lion with a gun.
- d. *p  **ye** ʈop k zmaray wuw ft 

Furthermore, the phrasal host must be stress-bearing; it must contain at least one primary accent. Unaccented constituents may not host 2P clitics, as the example in (2) illustrates.

- (2) a. [rata]_{PP} [pe]_{PP} [gand ]_V **de**
for-me by-him sew.PF.PST.3SG cl.2SG
You were having him sew (it) for me. (Hock 1996, p. 235)
- b. *[rata]_{PP} **de** [pe]_{PP} [gand ]_V

3. A Derivational Approach

A recent analysis of Pashto 2P clitics in Roberts (2000), articulated within the framework of the Minimalist Program (Chomsky, 1995), takes the position that (pronominal) 2P clitics are agreement morphemes. This assumption is based on the observation (also made in Babrakzai (1999)) that pronominal 2P clitics are in complementary distribution with verbal agreement morphology. That is, only ergative and accusative arguments may be cliticised. Nominative or absolutive arguments cannot. Roberts claims (following Sportiche (1996)) each clitic heads an agreement projection, whose specifier licenses a null pronominal argument. Under the analysis, 2P clitics are base-generated in high clausal position. The second position effect is derived through obliga-

tory syntactic movement. For example, the sentence in (3) receives the phrase structure analysis in FIGURE 2.

- (3) [spín]_{NP} **me** wənəlídə
 spin cl.1SG PF.NEG.SEE.PST.3SG.MASC
I didn't see Spin. (Roberts 2000, p. 81)

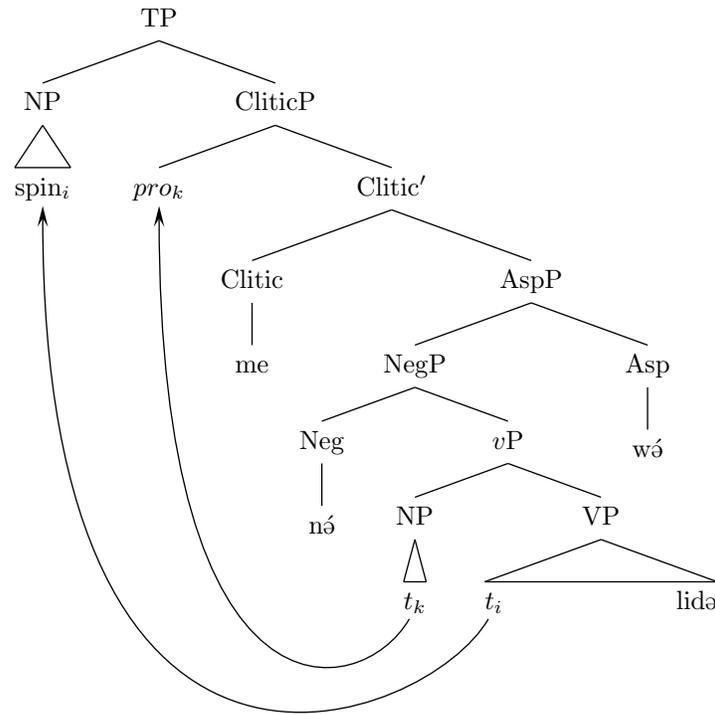


FIGURE 2 Phrase structure analysis of (3)

In the structure above, the pronominal clitic **me** is base-generated as the head of CliticP. The second position effect is derived by movement of the NP **Spin**, triggered by the EPP requirement that [Spec TP] be filled at Spell-Out (Roberts 2000, p. 83).

In Pro-drop clauses that contain only a verb and 2P clitics (henceforth, VC clauses), however, the EPP fails to trigger movement of the verb. As noted in Roberts (2000), a purely syntactic account incorrectly derives ungrammatical orderings. As the examples in (4) indicate, 2P clitics in VC clauses are hosted by the verb.

- (4) a. pórebdə **me**
 beat.IMPF.PST.3SG cl.1SG
 I was beating him. (Tegey 1977, p. 88)
- b. ***me** pórebdə
- c. likóm **ye**
 write.IMPF.PR.1SG cl.3SG
 I am writing it. (Babrakzai 1999, p. 93)
- d. ***ye** likóm

Roberts argues for a “last resort” application of Prosodic Inversion (Halpern, 1995, henceforth PI), which inverts 2P clitics with prosodic words on their right, to repair faulty syntactic derivations, as in FIGURE 3. At the post-syntactic level of PF, PI applies to repair syntactic

$$[\text{ye}] [\omega \text{ likóm}] \Rightarrow [\omega [\omega \text{ likóm}] \text{ye}]$$

FIGURE 3 Prosodic Inversion in (4c)

structures in which the 2P clitic is left without an appropriate prosodic host.

To summarise briefly, the approach proposed in Roberts (2000) argues that (pronominal) 2P clitics in Pashto are agreement morphemes base-generated in high clausal position. The second position placement of clitics within the clause is derived primarily in the syntax, with obligatory movement of an overt phrase to [Spec, TP], to satisfy the EPP. In VC clauses, the EPP does not force movement, and the offending structure is repaired post-syntactically by an application of PI which inverts 2P clitics with the nearest prosodic word.

3.1 Issues facing the proposal

The proposal in Roberts (2000) faces three main issues that I discuss in this section. First, the *clitics-as-agreement* hypothesis makes the prediction that 2P clitics should double overt pronominal arguments, which they do not. Secondly, the PI strategy that derives examples like (4a,c) encounters difficulty with several cases of apparent clitic infixation. Third, the intuition that 2P clitics are *prosodically* required to have hosts is lost by the primarily syntactic nature of the proposal. That is, the trigger for movement in general cases (the EPP) is a syntactic requirement and not a prosodic one.

The hypothesis that 2P clitics in Pashto are actually agreement morphemes suggests that they should cooccur with full nominal arguments. This prediction is not borne out in the data, as shown by (5).

- (5) a. zə kɪtəb əxləm
 I book buy.IMP.F.PR.1SG
I'm buying a book.
- b. *zə **me** kɪtəb wəxlə
 I cl.1SG book PF.buy.PST.3SG.MASC
I bought a book.

In (5a) the first person pronoun **zə** is doubled by first person inflection on the verb **əxləm**. The expectation is that the 2P clitic **me** should similarly double the pronoun, which (5b) shows not to be the case. Roberts (2000) addresses the issue and suggests that an explanation of the absence of 2P clitics doubling full NPs may be available if one assumes (following Jaeggli (1982)) that clitics absorb Case, rendering overt NPs without a means to satisfy the Case Filter.

A second issue facing the analysis in Roberts (2000) concerns several cases of apparent clitic infixation. Tegey (1977) identifies three classes of main verbs in Pashto that sometimes require 2P clitics to appear after the first accented syllable of the verb, rather than after the verb itself. Example (6) illustrates one such case, conditioned on the aspect of the verb **ʧəlwəhə**.

- (6) a. ʧəlwəhə́ **me**
 push.IMP.F.PST.3SG cl.1SG
I was pushing it. (Tegey 1977, p. 92)
- b. *ʧəl **me** wəhə́
- c. ʧəl **me** wəhə
 ? cl.1SG push.PF.PST.3SG
I pushed it. (Tegey 1977, p. 92)
- d. *ʧəlwəhə **me**

Examples (6c,d) show that when stress occurs on the first syllable of the verb (in the perfective aspect), 2P clitics obligatorily occur after the stressed syllable. The issue facing a PI account of clitic placement is how to appropriately define the prosodic element that inverts with the clitic. Cases like (6) suggest that a finer-grained notion of prosodic constituency than is traditionally assumed must be used to account for the Pashto facts.

A third issue concerns the conceptual awkwardness of the “syntax first—then prosody” derivation of the second position effect. The analysis seeks to derive second position by means of a purely syntactic movement (EPP). It is only if EPP movement fails to apply that prosodic

considerations are brought to bear on clitic placement. This seems to conflict with the intuition that clitics are prosodically dependent and “lean” on some adjoining host element. Note that the *clitics-as-agreement* approach does away with clitics as a natural class. However, the issue concerns 2P clitics empirically, regardless of what theoretical status they have. The division of labour, so to speak, for ensuring clitics have hosts does not match well with general observations about their prosodic dependencies.

To summarise briefly, the approach taken in Roberts (2000) faces three issues worth pointing out. First, identifying clitics with agreement morphemes makes an incorrect empirical prediction about their behaviour with respect to clitic doubling. Secondly, it is unclear how the PI strategy is to be constrained such that examples like (4a) and (6c) are uniformly predicted. Finally, the split nature of how clitic placement is derived clashes with the intuition that clitics have prosodic demands of their hosts. The data suggest a unified treatment of clitic positioning would be more appropriate, which I explore in the following sections.

4. A Domain-Based Approach

4.1 Architecture

HPSG is a constraint-based theory of grammatical competence that relies on two fundamental components: a highly structured representation of grammatical categories, encoded as typed feature structures, and a set of descriptive constraints on types, which restricts the expressions admitted as part of a given natural language (Levine and Meurers, to appear). The set of descriptive constraints are declarative and unordered, making HPSG non-derivational. Formally, an HPSG grammar is given by a set of *type constraints* and a *signature*. The signature consists minimally of an enumeration of the set of grammatical *types* (also called *sorts*), as well as a statement of which features are appropriate for each type and a statement of what type of value is appropriate for each feature (Ginzburg and Sag 2001, p. 17). The basic unit of currency in HPSG is the *sign*, which encodes linguistic information such as the syntactic/semantic, phonological and contextual properties of grammatical objects. Signs are represented as Attribute-Value Matrices (AVMs) of the sort in FIGURE 4.¹ Both word-level and phrase-level objects are analysed in terms of *signs*. Syntactic/semantic information

¹I will employ the following notational conveniences throughout this paper: crucial aspects of AVMs will be highlighted **as so**. Coindexed boxed numerals (like $\boxed{1}$) indicate structure-sharing of feature content. For space reasons, features and values will be abbreviated where possible.

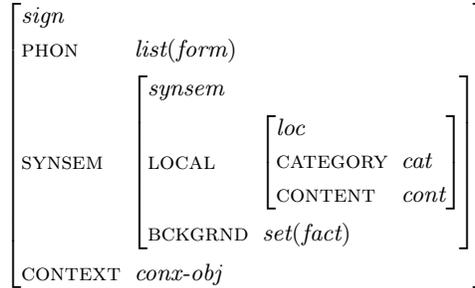


FIGURE 4 (Part of) the structure of *sign*

is encoded on the feature SYNSEM, and phonological/prosodic properties are encoded on the feature PHON(OLGY). The type *synsem* may be viewed as an analogue to traditional syntactic category representations (e.g., N, V, Adj). Phrasal *signs* are licensed by type constraints on well-formed objects. For example, the Generalised Head Feature Principle (GHFP, see (7) below) licenses instances of *h(ea)d(ed)-ph(rase)* such that the SYNSEM value of the mother is structure-shared (i.e., token-identical) with the SYNSEM value of exactly one head daughter. This is analogous to the ‘X’ identity condition of \bar{X} Theory (Ginzburg and Sag 2001, p. 33).

(7) a. *hd-ph*:

$$\left[\text{SYNSEM} \quad / \square \right] \rightarrow \dots \mathbf{H} \left[\text{SYNSEM} \quad / \square \right] \dots$$

a'. Headed phrases contain exactly one head daughter.

Additional type constraints that will be relevant to the present discussion are given explicitly in (8) below.

(8) a. *hd-comp-ph*:

$$\left[\right] \rightarrow \mathbf{H} \left[\begin{array}{l} \textit{word} \\ \text{COMPS} \quad \textit{nelist}(\square \oplus \textit{list}) \end{array} \right], \square$$

a'. Head-complement phrases (e.g., VP) contain a lexical head and its complements.

b. *hd-subj-ph*:

$$\left[\text{SUBJ} \quad \langle \rangle \right] \rightarrow \left[\text{SYNSEM} \quad \square \right], \mathbf{H} \left[\begin{array}{l} \text{SPR} \quad \langle \rangle \\ \text{SUBJ} \quad \langle \square \rangle \end{array} \right]$$

b'. Head-subject phrases (e.g., S) contain a head daughter and exactly one subject.

To illustrate with a small example, the English sentence **Randi likes kittens** receives the phrase structure analysis in FIGURE 5. The verb

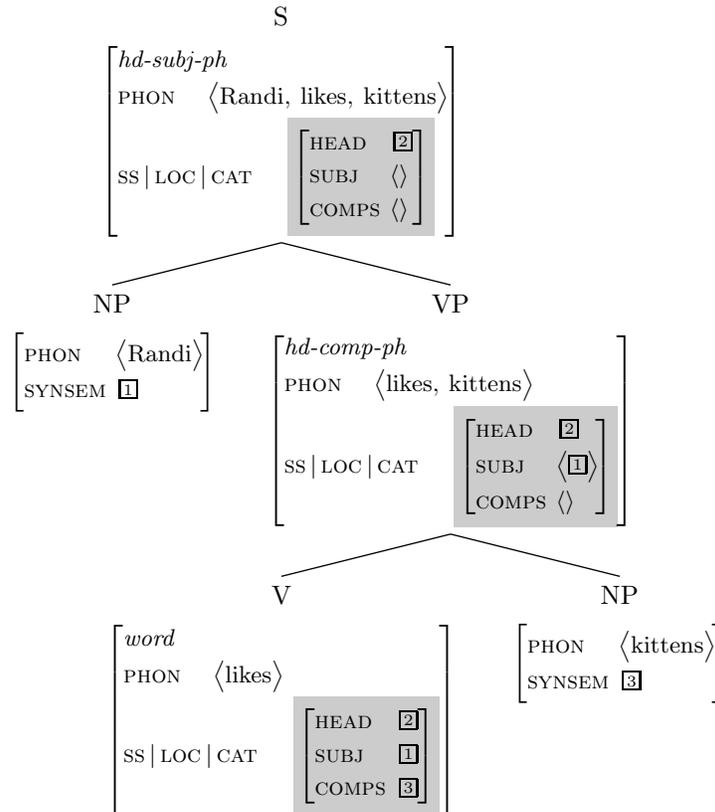


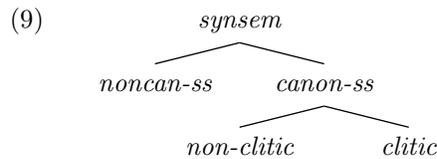
FIGURE 5 (Partial) syntactic analysis of **Randi likes kittens**

likes subcategorises for two arguments: the object NP **kittens** (by way of the feature COMPS) and the subject **Randi** (by way of SUBJ). The VP (an instance of *hd-comp-ph*) structure shares its HEAD features with those of its head daughter (V). Note that at the VP level, the COMPS value of the main verb has been saturated. The clause (an instance of *hd-subj-ph*) is (partially) licensed by the type constraint in (8b). The VP combines with the subject **Randi**. Again, the SUBJ feature on the mother node bears an empty list, corresponding with argument saturation. A grammatical analysis of the sentence emerges through the interaction of lexical entries for the terminal elements, and type

constraints that mediate syntactic structure.

4.2 The syntax of 2P clitics

The intuition I wish to pursue in this section (originally proposed in Kupść (2000) for Polish clitics) is that the clitic *vs.* non-clitic distinction is orthogonal to the syntactic properties of lexical *signs*. That is, the distinction between clitics and non-clitics is one that runs parallel to the purely syntactic aspects of lexical items. Hence, an (*e.g.*, pronominal) clitic is the same syntactic category as its plain-word analogue, but represents a different class of *synsem* object. Following Kupść (2000), I adopt the partition of the type *canonical-synsem* into *clitic* and *non-clitic* subtypes, as shown in (9).



Setting aside for the moment the featural differences between *clitic* and *non-clitic* types², the approach I take here treats clitics identically with their non-clitic counterparts from a syntactic point of view. The same constraints that license overt pronominal arguments in Pashto clauses will also license pronominal 2P clitics, and similarly for modal and adverbial clitics. To illustrate, the clitic **me** in (3) has the lexical entry in FIGURE 6 ((3) repeated here as (10)).

- (10) spín **me** wənəlīdə
 spin cl.1SG PF.NEG.see.PST.3SG.MASC
I didn't see Spin.

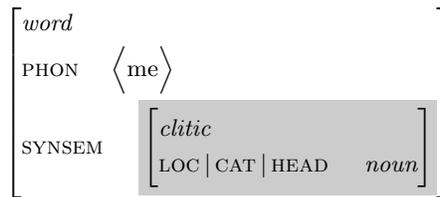


FIGURE 6 (Partial) lexical entry for **me**

²Simply, the distinction between *clitic* and *non-clitic* SYNSEM types is based on prosodic deficiency. Clitics subcategorise for a prosodic host, while non-clitics do not. Interested parties should consult Klein (2000) for a constraint-based approach to prosody.

4.3 Word order domains

Domain-based approaches to linear ordering phenomena within the HPSG literature (Reape, 1994, Kathol, 1995, Penn, 1999, among others) represent one of several strands of research that depart from traditional notions regarding the relationship between the syntactic structure and linear organisation of strings (see Kathol (1995) for further discussion of the history of domain-based approaches to word order phenomena). The intuition behind word order domains is that constraints on hierarchical structure are related to, but formally distinct from those on linear realisation, a claim made as early as Curry’s (1961) distinction between *tectogrammatical* structure, where meanings are constructed compositionally, and *phenogrammatical* structure, where constraints on the surface realisation of a string are resolved (see also Dowty (1982, 1996)).

In HPSG approaches to word order domains, the ordering of elements that make up the terminal yield of a *sign* (word order) is encoded on a *sign* level attribute $\text{DOM}(\text{AIN})$. A general constraint requires the phonological yield of a sign to be equivalent to the concatenation of the PHON values of objects on its DOM list, as shown in FIGURE 8. The

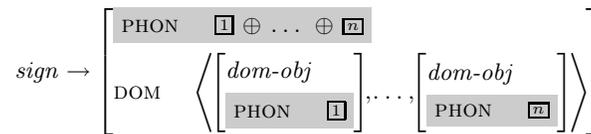


FIGURE 8 Constraint on the phonological yield of *sign* (Reape, 1993)

constraint in FIGURE 8 simply requires the PHON value of a sign to correspond with the PHON values of each of its domain elements.

The analysis I propose here also makes use of the notion of *topological fields*, as used in Kathol (1995). Briefly, topological fields are position classes that model the linear distributional properties of syntactic elements, without making claims about their hierarchical organisation.³ The intuition is that word order patterns of a language can be described in terms of membership in (linearly) ordered fields. Kathol (1995) makes formal use of such fields to account for word order effects in German clauses.

The intuition behind the analysis of 2P clitic placement in Pashto clauses is that the linear ordering of a grammatical object is licensed in

³Interested parties should consult Höhle (1986) for a review of topological fields theory.

tandem with its hierarchical structure. Constraints on word order and those on hierarchical organisation operate in parallel. Additionally, the surface constraints on linear order are argued to result from interactions between topological fields, syntactic constraints, and prosodic requirements.

As a starting point, I assume the topological model for simple Pashto clauses in FIGURE 9. In the proposed topology, elements that host 2P

	Pre-clitic (<i>pre-cf</i>)	Clitic (<i>cf</i>)	Middle (<i>mf</i>)	Verbal (<i>vf</i>)
a.			zə kɾɪɒb I book	axləm buy.IMP.F.PR.1SG
b.	spín spín	me cl.1SG		wónəlídə PF.NEG.see.PST.3SG
c.	likám write.IMP.F.PR.1SG	ye cl.3SG		

FIGURE 9 Topological spaces in simple Pashto clauses

clitics instantiate the Pre-clitic (*pre-cf*) field, while 2P clitics occur in the Clitic (*cf*) field. The Middle (*mf*) field is the default field assignment for the clause, and verbal elements instantiate the Verbal (*vf*) field, anchoring the edge of the clause. This division of fields is motivated in part by grammatical variations between OSV and SOV word order, illustrated in (11).⁴

- (11) a. spi pɪʃo xog kɾa
dog cat hurt do.PST.PF.3SG.FEM
The dog hurt the cat. (Roberts 2000, p. 12)
- b. pɪʃo spi xog kɾa
cat dog hurt do.PST.PF.3SG.FEM
The dog hurt the cat.

Following Kathol (1995), I formalise the notion of topological fields as a type *topo(logical)-field*, with the hierarchy in FIGURE 10. A few important things to note about the hierarchy of topological field types; HPSG feature descriptions are maximal, therefore all field assignments must be resolved. The nonmaximal types I assume are *verbal* and *nonverbal*. The *verbal* type dominates both *pre-cf* and *vf* assignments. Intuitively,

⁴OSV word order in Pashto clauses is more complicated than I wish to address in this paper. Specifically, OSV ordering is permissible when inflectional morphology on the verb disambiguates grammatical role. In the absence of suitable inflection (i.e., in present tense clauses), OSV ordering requires a heavy pause after the direct object.

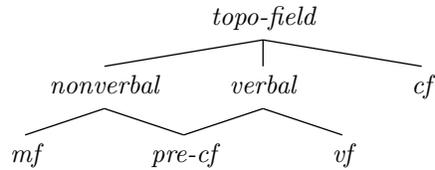


FIGURE 10 Hierarchy of topological field types

verbal lexical items can either occur clause finally (as in (1)) or can host clitics (as in (4)). Similarly, *nonverbal* field assignments can resolve to either *mf* (general case) or *pre-cf* (when hosting 2P clitics).

As mentioned earlier, the constraints on 2P clitic placement arise from the interaction of relatively simple constraints in different areas of the grammar. I make use of two constraints on topological fields, given in (12).

- (12) a. **Linear Precedence:**
 $pre-cf \prec cf \prec mf \prec vf$
 a'. Topological fields occur in a fixed order.
- b. **Field Uniqueness:**
 $pre-cf \prec pre-cf$
 b'. The *pre-cf* field must be unique (*i.e.*, contain a single member).

Linear Precedence requires typed topological fields to occur in a specific order. Field Uniqueness constraint ensures that in any Pashto clause, only one object may instantiate the *pre-cf* field.

Again, following Kathol (1995) I assume that syntactic constraints on field assignments may be either lexical or constructional. That is, lexical constraints condition field assignments based on lexical class, while constructional constraints condition field assignments in phrasal *signs*. I propose that *cf* and *verbal* field assignments are lexical, applying to 2P clitics and verbs, respectively. The relevant constraints are given in (13).

- (13) a. $clitic \Rightarrow \left[\text{DOM} \left\langle \left[\begin{array}{l} cf \\ PHON \dots \end{array} \right] \right\rangle \right]$
- a'. 2P clitics are typed *cf*.
- b. $verb \Rightarrow \left[\text{DOM} \left\langle \left[\begin{array}{l} verbal \\ PHON \dots \end{array} \right] \right\rangle \right]$
- b'. Verbs are typed *verbal*.

Since the domain of the present analysis is restricted to simple Pashto clauses, a constraint on field assignment in headed phrases (*hd-ph*) is given in (14).

- (14) **Domain Condition:** (Informal version)
 In all instances of headed phrase (*hd-ph*) such that:
 (a) the head daughter is typed *verbal*
 (b) all non-head daughters contain *non-clitic* SYNSEM values
 the head daughter is typed *vf* and all non-head daughters are typed *nonverbal*.

The constraint above targets phrases headed by a *verbal* element and containing no 2P clitics. Such phrases are constrained to resolve the head to the *vf* field assignment and all non-head daughters to the *non-verbal* assignment. The constraint has the effect of ensuring that verbs occur on the right edge of the clause, in all cases where there is overt phonological material to host 2P clitics.

Returning to the distinction between *clitic* and *non-clitic* synsems, I propose a prosodic constraint on all 2P clitics requiring them to have phonological hosts. Clitics are distinguished from non-clitics with respect to prosodic dependency. Intuitively, instantiation of the *pre-cf* field can be seen as a reflex of this requirement. That is, the second position effect under the current analysis receives a unified explanation in terms of prosodic subcategorisation.

- (15) **Clitic Condition:** (Informal version)
 2P clitics subcategorise *full* prosodic structures to their immediate left.⁵

⁵Here, *full* denotes a type of prosodic structure in Klein (2000) that subsumes both prosodic words and larger metrical structures. In the interest of space, I refrain from a full discussion of the analysis of prosodic structure in Pashto. The crucial point I wish to make here is only that prosodic dependencies can be modeled as well formedness conditions on structures that contain clitics. The current analysis is compatible with a number of ways of modeling prosodic dependency in HPSG.

To illustrate how the analysis here accounts for the various 2P clitic placement patterns in Pashto, I present three representative examples and discuss how the analysis predicts each. First, consider the now-familiar example in (3) and (10) (in its third incarnation, as (16)). Under the present approach, the sentence receives the domain ordering analysis in FIGURE 11.

- (16) *spín me wənəlīdə*
 spin cl.1SG PF.NEG.see.PST.3SG.MASC
I didn't see Spin.

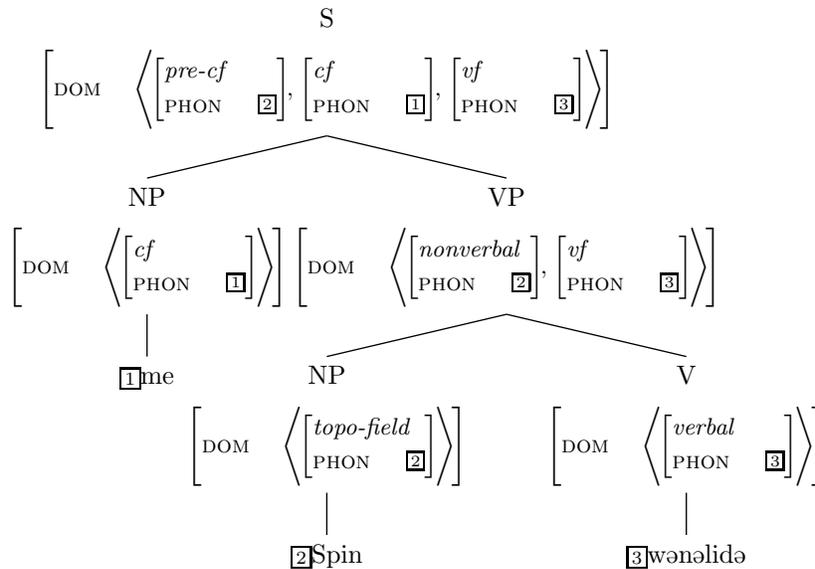


FIGURE 11 Ordering analysis of (16)

The lexical constraints on field assignment in (13) ensure that the verb *wənəlīdə* is typed *verbal* and the 2P clitic *me* is typed *cf*. This means the verb must resolve to either *vf* or *pre-cf* at the level of the clause (see FIGURE 10). At the level of the VP, the Domain Condition applies, and the verb is resolved to *vf*, while the NP *Spin* resolves to *nonverbal*. Finally, at the S-level, the Clitic Condition forces a resolution of the object NP to *pre-cf*, satisfying the prosodic dependency of the clitic. Field ordering is enforced throughout by Linear Precedence and Field Uniqueness.⁶

⁶While it seems intuitive to think of field resolution in terms of derivations, there

Consider now examples such as (4), which are accounted for in Roberts (2000) by means of a post-syntactic PI movement. The example in (4c) (repeated here as (17)) receives the domain ordering analysis in FIGURE 12.

- (17) **likəm** **ye**
 write.IMPF.PR.3SG cl.3SG
I am writing it.

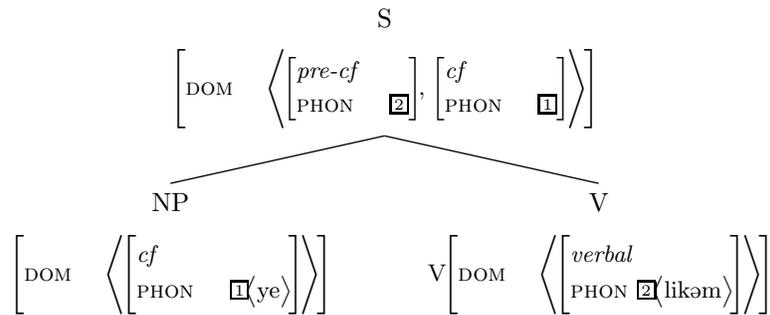


FIGURE 12 Ordering analysis of (17)

As in the previous example, the verb **likəm** and 2P clitic **ye** are typed *verbal* and *cf* respectively, by virtue of the lexical constraints in (13). The resulting phrase is not subject to the Domain Condition, as it contains a *clitic* typed SYNSEM value. At the topmost level, the Clitic Condition compels a resolution of the *verbal* domain object to *pre-cf*, to ensure the clitic has a host. As before, the ordering of fields and uniqueness of the *pre-cf* field are mediated by the constraints in (12). In the analysis proposed here, the difference between cases like (1) and (4) has to do with application of the Domain Condition. That the 2P clitic has a host in both cases receives a uniform explanation in terms of prosodic dependency.

Finally, I address examples like (6), in which the 2P clitic appears to split lexical items. In order to account for these data, I make the additional assumption (following Crysmann (2000)) that word-level entities may contribute more than one domain object.⁷ That is, verbs like

is nothing derivational about the current account. Resolutions of topological fields to any other assignments render the string ungrammatical. That is, there is exactly one assignment that corresponds with a grammatical clause. For expository clarity, explanations of domain analyses may have this flavour.

⁷The analysis in Crysmann (2000) targets subject agreement markers in Udi, which display nontrivial similarities to 2P clitic placement in Pashto. Specifically,

not apply (the phrase contains a *non-clitic* valued SYNSEM feature). The Clitic Condition forces a resolution of the domain object associated with **tə**l to the *pre-cf* field. Field Uniqueness ensures that only one *pre-cf* typed element occurs in the clause. The remaining element must resolve to *vf* in order for the clause to be licensed.

To summarise, in the analysis of their syntactic licensing, 2P clitics are treated as separate classes of *synsem*, capturing the intuition that they are different from their non-clitic analogues in some respects (prosodic deficiency), and not in others (occur as arguments to verbs, etc.). The analysis requires nothing new to be said about how syntactic objects containing 2P clitics are licensed. Furthermore, since pronominal clitics are treated as arguments, the analysis avoids the prediction that they should double overt NPs. The analysis of Pashto clause ordering assumes a topology for simple clauses in which the verb anchors the right edge of the clause. Clitic hosts instantiate the *pre-cf* field, clitics occur in the *cf* field, and the default field assignment is *mf*. The intuition pursued by the analysis is that word order effects are constrained in parallel with syntactic structure. Word order is argued to result from the interaction of multiple grammatical components, including syntax and prosody. Assuming the topological model of simple Pashto clauses in FIGURE 9, and the hierarchy of topological types in FIGURE 10, the proposed analysis accounts for 2P clitic placement through the interaction of simple constraints. Topological constraints mediate field ordering and uniqueness. Lexical constraints specify field assignments that are present in the lexical entries of both 2P clitics and verbs. The Domain Condition mediates field assignment in headed phrases, and the Clitic Condition is a prosodic requirement on 2P clitics that is responsible for “deriving” the second position effect. The immediate benefits of the current proposal include a prediction of the absence of clitic doubling, a unified treatment of the second position effect in terms of prosodic subcategorisation, and an account of apparent clitic infixation that avoids the complications of a PI-based account.⁸

5. Remarks and Conclusions

Some issues remain for the analysis presented in this paper that merit brief discussion here. First, the analysis does not explicitly address the ban on pronominal 2P clitics from appearing as objects of adpositions. Moreover, this seems to be a general ban cross-linguistically. A solution

⁸The Clitic Condition merely requires the host of a 2P clitic to be of the prosodic type *full*. This underspecification allows for the clitic’s host to range from larger metrical structures to prosodic words, and in the case of examples like (6), sublexical prosodic units.

along the lines of prosodic incompatibility is suggested by the proposal here. Assuming the Clitic Condition requires 2P clitics to have suitable prosodic hosts, the ban on such clitics appearing in PPs receives an explanation if one argues that adpositional elements in Pashto are ill-suited for this role. That is, elements of category P do not constitute *full* prosodic objects. A second issue that remains to be addressed concerns the correspondence between prosodic and syntactic structure. The current account requires 2P clitics to be hosted by a *full* prosodic element. In cases like (1), the correspondence of an S-level XP with a *full* prosodic object must be investigated. I argue above that the analysis here assumes a constraint-based phonology for Pashto clauses as in Klein (2000). The constraints responsible for the mapping of syntactic structure to prosodic structure need to be fully fleshed out; a matter I reluctantly leave for future research.

In conclusion, this paper presents an analysis of 2P clitics in Pashto that makes use of word order domains as a means to constrain the linear realisation of clauses containing them. The approach proposed here makes a formal distinction between constraints on hierarchical structure, and those on linear organisation. The obvious benefits of such a strategy are a simplified account of the syntax of 2P clitics, and an account of their placement that makes reference to both syntactic and prosodic information. A derivational account in Roberts (2000) was shown to face certain complications. Specifically, the *clitics-as-agreement* hypothesis does not make obvious gains, as it must be supplemented with a prosodic repair strategy (PI) to account for the basic alternations, and it makes an incorrect empirical prediction regarding clitic doubling. Additionally, the derivational account misses the intuition that the second position effect in Pashto clauses is the result of a prosodic dependency, by dividing the mechanisms that account for 2P clitic placement. The analysis presented here offers a unified treatment of the second position effect, and simplifies the syntactic assumptions required to account for 2P clitic licensing. The domain ordering component of the analysis was shown to follow straightforwardly through the interaction of different constraints, each relatively simple in its formulation. The collaborative effect correctly predicts 2P clitic placement in a variety of different cases, including phrasal hosts (see 1), verbal hosts (see 4) and apparent infixation (see 6). An interesting result of the proposal here concerns the infixation cases and the principle of Lexical Integrity (Bresnan and Mchombo, 1995). At first glance, it appears as though Pashto does violence to the claim that syntactic processes should not make reference to the internal composition of lexical items. However, under the current approach, the syntactic component of the

grammar is not at fault, so to speak. Crucially, I have argued that 2P clitic placement is not a purely syntactic process, but rather one that results from the interaction of syntax, prosody and word order constraints. The account here is strictly compatible with the notion that *syntactic* processes do not reference the internal makeup of words. As a final thought, it is hoped that the research here stimulates further investigation of the role that domain ordering analyses might play in accounting for phenomena that appear to be sensitive to multiple components of the grammar.

Acknowledgments

I would like to thank (in no particular order) Mohammad Dost, Shahab Khan, Andreas Kathol, Geoff Pullum, Judith Aissen, Sandy Chung, Peter Sells, and the audience at TLS 9 for valuable comments and insights. The work presented here is greatly improved as a result of their input. All remaining errors and omissions have been placed deliberately by the author in order to encourage future research in this area.

References

- Babrakzai, Farooq. 1999. *Topics in Pashto Syntax*. Ph.D. thesis, University of Hawai'i at Manoa.
- Bresnan, Joan and Sam Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13:181–254.
- Chomsky, Noam. 1995. *The Minimalist Program*, chap. 3, pages 167–217. MIT Press.
- Crysmann, Berthold. 2000. On the placement and morphology of Udi subject agreement. paper presented at the 7th International Conference on Head-driven Phrase Structure Grammar, Berkeley.
- Curry, Haskell B. 1961. Some logical aspects of grammatical structure. In *Structure of Language and its Mathematical Aspects: Proceedings of the Twelfth Symposium in Applied Mathematics*, pages 56–68. American Mathematical Society.
- Dowty, David. 1982. Grammatical relations and Montague grammar. In P. Jacobson and G. K. Pullum, eds., *The Nature of Syntactic Representation*, pages 79–130. Dordrecht: D. Reidel.
- Dowty, David. 1996. Towards a minimalist theory of syntactic structure. In H. Bunt and A. van Horck, eds., *Discontinuous Constituency*, pages 11–62. Berlin/New York: Mouton de Gruyter.
- Ginzburg, Jonathan and Ivan A. Sag. 2001. *Interrogative Investigations: the form, meaning, and use of English Interrogatives*. Stanford, California: CSLI Publications.

- Halpern, Aaron. 1995. *On the Placement and Morphology of Clitics*. Stanford: CSLI.
- Hock, Hans Henrich. 1996. Who's on first? Toward a prosodic account of P2 clitics. In A. L. Halpern and A. M. Zwicky, eds., *Approaching Second: Second Position Clitics and Related Phenomena*, pages 199–270. CSLI Publications.
- Höhle, Tilman. 1986. Der Begriff “Mittelfeld”, Anmerkungen über die Theorie der topologischen Felder. In A. Schoöne, ed., *Kontroversen, alte und neue: Akten des 7. Internationalen Germanisten-Kongresses, Göttingen 1985, Band 3*, pages 329–340. Tübingen: Niemeyer.
- Jaeggli, Osvaldo. 1982. *Topics in Romance Syntax*. Dordrecht: Foris Publications.
- Kathol, Andreas. 1995. *Linearization-Based German Syntax*. Ph.D. thesis, Ohio State University.
- Klein, Ewan. 2000. Prosodic constituency in HPSG. In R. Cann, C. Grover, and P. Miller, eds., *Grammatical Interfaces in HPSG*, pages 169–200. CSLI Publications.
- Kupść, Anna. 2000. *An HPSG Grammar of Polish Clitics*. Ph.D. thesis, Institute of Computer Science, Polish Academy of Sciences and Université Paris 7, Warszawa, Poland.
- Levine, Robert D. and Walt Detmar Meurers. to appear. Head-driven phrase structure grammar: Linguistic approach, formal foundations, and computational realization. In K. Brown, ed., *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 2nd edn.
- Penn, Gerald. 1999. An RSRL formalization of Serbo-Croatian second position clitic placement. In V. Kordoni, ed., *Tübingen Studies in Head-Driven Phrase Structure Grammar*, no. 132 in Arbeitsberichte des SFB 340, pages 177–197. Tübingen: Universität Tübingen.
- Pollard, Carl J. and Ivan A. Sag. 1987. *Information-based Syntax and Semantics, Vol. 1*. No. 13 in CSLI Lecture Notes. Stanford University: CSLI Publications. Distributed by University of Chicago Press.
- Pollard, Carl J. and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Reape, Mike. 1993. *A Formal Theory of Word Order: A Case Study in West Germanic*. Ph.D. thesis, University of Edinburgh.
- Reape, Mike. 1994. Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. J. Pollard, eds., *German in Head-Driven Phrase Structure Grammar*, no. 46 in CSLI Lecture Notes, pages 151–197. Stanford University: CSLI Publications.
- Roberts, Taylor. 2000. *Clitics and Agreement*. Ph.D. thesis, Massachusetts Institute of Technology.
- Sag, Ivan A. and Thomas Wasow. 1999. *Syntactic Theory: A Formal Introduction*. Stanford: Center for the Study of Language and Information. Japanese edition (two volumes) – translated and edited by Takao Gunji and Yasunari Harada, appeared in 2001.

- Sportiche, Dominique. 1996. Clitic constructions. In J. Rooryck and L. Zaring, eds., *Phrase Structure and the Lexicon*, pages 213–276. Dordrecht: Kluwer Academic Publishers.
- Tegey, Habibullah. 1977. *The grammar of clitics: evidence from Pashto and other languages*. Ph.D. thesis, University of Illinois at Urbana-Champaign.
- Zwicky, Arnold and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59:502–513.

***-ing* as an Agreement Marker in African American English: Implications for Acquisition¹**

LISA GREEN

1 Introduction

This paper focuses on *-ing* in aspectual constructions in African American English (AAE) and argues that it functions as a morphological agreement marker that is required by the feature [HABITUAL], not as a progressive marker. Section 2 presents a general overview of the distribution of the forms of auxiliary *be*, which will be compared to aspectual (habitual) *be* in a later section. Section 3 considers the remote past marker *BIN* and suggests that *-ing* in *BIN V-ing* sequences is linked to the ambiguity in those constructions. *V-ing* in *BIN V-ing* constructions can have a progressive (or state) reading, or it can have a habitual reading. Section 4 presents an overview of aspectual *be* constructions, which are distinguished from progressive constructions and simple tense generics. Section 5 presents an analysis of *-ing* as an agreement marker that is required to occur on verbs in habitual remote past *BIN* and aspectual *be* constructions. The final section of the paper presents data from child comprehension tasks, which raise questions about the extent to which developing AAE speakers understand the habitual

¹Part of this research was supported by a grant from the National Science Foundation (BCS-0214388). I would also like to thank Carlota Smith, Kendra Bradford, Rebecca Quigley, Nikki Seifert, and Jessica White for their contributions.

interpretation associated with aspectual markers *be* and *BIN*. In addition, the data raise questions about the extent to which children begin to associate *-ing* in aspectual *be* and remote past *BIN* *V-ing* constructions with the feature [HABITUAL] as they develop the AAE tense-aspect system.

2 Distribution of Forms of *Be_{aux}*

There are at least two BEs in AAE: the auxiliary (and copula) and aspectual *be*. The auxiliary *be* (*be_{aux}*) in AAE can be spelled out in its overt (contracted or full) forms, or it can be null (Labov 1969; Baugh 1980; Rickford, Ball, Blake, Jackson, and Martin 1991; Déchaine 1993; Green 1993; Walker 2000). Examples of the overt *be_{aux}* are in (1):²

- (1) a. Dee *be_{auxi}* + [EMPH] *t_i* running → Dee IS running.
- b. Dee *be_{auxi}* + [PAST] *t_i* running → Dee was running.
- c. Q*be_{auxi}* + [-PAST] *t_i* → IS Dee running?
- d. Dee *be_{auxi}* + [PAST] *t*+NEG_{*i*} running → Dee wasn't running.
- e. 1st singular [-PAST]: I'm running.
- f. 3rd singular neuter [-PAST]: It's running.

The data in (1) show that auxiliary *be* occurs in its overt form when it hosts an emphatic morpheme (1a), past tense (1b), question morpheme (*Q*) (1c), and past tense and negation morphemes (1d). In addition, *be_{aux}* surfaces in the environment of first person singular non-past and third person singular non-past neuter pronouns. Null (\emptyset) *be_{aux}* occurs when it is not required to host past tense or a *Q* morpheme, as illustrated in (2). The *Q* morpheme may signal question intonation and interrogative force and may be different from the one in (1c), which requires a host.

- (2) a. Dee \emptyset *be_{auxi}* + [-PAST] *t_i* running → Dee running.
- b. Dee \emptyset *be_{auxi}* + [-PAST] *t_i*+NEG_{*i*} running → Dee not running.
- c. Q Dee \emptyset *be_{auxi}* + [-PAST] *t_i* running → Dee running?

3 *-ing* and the Progressive

Be_{aux} (overt or \emptyset) *V-ing* marks the progressive in AAE just as it does in other varieties of English.

- (3) a. Dee running. ('Dee is running')

²The copula can also be null in AAE.

- b. Dee was running.
- c. Dee was sweeping the floor when the phone rang.

As in mainstream American English, stative verbs do not generally occur in the progressive in AAE (4a); however, there are contexts in which this restriction is relaxed (4b).

- (4) a. *Dee is/Ø knowing the answers.
- b. Dee is looking more and more like her mother these days.

V-ing also occurs as the predicate in *BIN* constructions. The remote past marker *BIN*, which is stressed, situates an eventuality or part of it in the remote past (Green 1998). *BIN V-ing* constructions are ambiguous between two readings that are labeled as *BIN_{STAT}* and *BIN_{HAB}*.

BIN_{STAT} situates the initial point of a state in the remote past. These *BIN* constructions are similar to progressives in that they also present “stable situations” (Smith 1997: 84).

- (5) a. Dee BIN running.
‘Dee has been running for a long time’
- b. Dee BIN knowing Swahili.
‘Dee has known Swahili for a long time’

These *BIN* constructions can be represented in a Parsons-type (1990) analysis in which events are argued to underlie the eventualities to which sentences refer:

- (6) a. Dee BIN running.
- b. $(\exists I)[\text{long}(I) \ \& \ \text{Beg}(I) < \text{now} \ \& \ (\exists e)(\exists s)[\text{running}(e) \ \& \ \text{theme}(e, \text{Dee}) \ \& \ \text{IP state}(e,s) \ \& \ \text{Hold}(s, I)]$

The representation in (6b) indicates that some long interval begins before now, and the running event holds throughout that interval.

BIN_{HAB} situates the initial point of a habit in the remote past:

- (7) a. Dee BIN running for 30 minutes.
‘For a long time Dee has had the habit of running for thirty minute stretches’

In (7) temporal modification is restricted to the verb. Modification is of periods of shorter instantiations of eventualities, running events expressed by the verb and that constitute the habit. Temporal adverbials cannot modify

BIN or the length of the long interval. The sentence in (7a) can be represented as in (7b):

- b. $(\exists I)$ [long(I) & Beg(I) now (i) $i \in I$ & HAB_i [for 30 minutes, i]
 $(\exists e)(\exists s)$ [running(e) & theme(e, Dee) & IP state(e, s) & Hold(s, I)]]

Given the restriction on modification, (7a) cannot have the *BIN_{STAT}* reading that means that Dee's running started 30 minutes ago and has held throughout the 30-minute interval. In general, non-stative verbs can have both the *BIN_{STAT}* and *BIN_{HAB}* readings, and, in some contexts, stative verbs can also have both readings. This was also shown to be the case with *be_{aux}* V-*ing* progressives. (See 4a.) A summary of *BIN* readings is given below:

a. \checkmark BIN _{STAT} , \checkmark BIN _{HAB}	Dee BIN running. (5a, 6a)
b. \checkmark BIN _{STAT} , \checkmark BIN _{HAB}	Dee BIN looking like her mother. BIN _{STAT} 'Dee has looked like her mother for a long time' BIN _{HAB} 'For a long time, Dee has looked like her mother for periods of time'
c. \checkmark BIN _{STAT} , \checkmark BIN _{HAB}	Dee BIN having that car. BIN _{STAT} 'Dee has had that car for a long time' BIN _{HAB} 'For a long time, Dee has used that car from time to time'
d. \checkmark BIN _{STAT} , ?#BIN _{HAB}	Dee BIN knowing how to fix washing machines.
e. \checkmark BIN _{STAT} , #BIN _{HAB}	Dee BIN knowing Swahili.

Table 1 Summary of *BIN* Readings

Both the states indicated by the predicates *looking* (Table 1, b) and *having* (Table 1, c) can have an event reading in habitual contexts. However, the state indicated by the verb *know* is somewhat more resistant to an event reading, as shown in (Table 1, d) and (Table 1, e). While it may be slightly possible to force the *BIN_{HAB}* reading of *know how to fix washing machines*, it is not at all possible to get this reading for *knowing Swahili*, which shows that the type of predicate in the *BIN* construction has some effect on the interpretation.

4 -*ing* and Habitual

In addition to occurring in progressive contexts, V-*ing* also occurs in habitual *be* constructions, which can be distinguished from progressive contexts.

In habitual constructions, habitual or aspectual *be* (*be_{asp}*) indicates that an eventuality recurs:

- (8) Dee be running.
'Dee usually runs/is usually running'

Habitual *be* constructions are similar to simple tense generics, but they can be distinguished from generics in that *be_{asp}* constructions are not ambiguous between habitual/generic and ability readings, but simple tense generics are:

- (9) a. Bruce work on old Thunderbirds.³
b. Bruce be working on old Thunderbirds.
'Bruce works on old Thunderbirds from time to time'

While the sentence in (9a) can have the reading Bruce will work on Thunderbirds or has the ability to work on old Thunderbirds although he may not have had the opportunity to do so, or it can mean that he actually works on old Thunderbirds from time to time.⁴ The sentence in (9b) can only have the universal reading, in which Bruce does indeed work on old Thunderbirds from time to time; it cannot have the ability reading in which Bruce can work on old Thunderbirds but has never actually worked on one. The *be_{asp}* construction such as that in (9b) has as its core reading habitual, but it is expressed with *be V-ing*, not simple tense. The *be V-ing* construction is similar in morphological form to the progressive, and it also has an in-progress reading:

- (10) Dee be riding her bike when the phone ring.
a. in-progress reading: Dee's bike riding is in progress when the phone rings.

However, the sentence in (10) also has the closed reading (Smith 1997), in which the bike riding event is not in progress when the phone rings:

- b. closed reading: Dee's bike riding begins after the phone rings.

The reading in (10b) clearly distinguishes the *be_{asp}* construction from the progressive. The habitual *be* construction differs from the progressive in one

³Person and number agreement marking in AAE is variable at best. In most cases, there is no overt person and number agreement marking in 3rd person singular contexts.

⁴The characterization of the generic reading as an ability reading is based on the notion of capacity reading in Schubert and Pelletier (1989).

additional way. As shown in (11a), *be_{asp} V-ing* is compatible with both states and events; however, as shown in (11b), progressive *be_{aux} V-ing* is not compatible with stative verbs:⁵

- (11) a. Dee be knowing how to fix washing machines.
 Literally: Dee usually shows that she knows how to fix washing machines (by repairing broken machines, giving advice about which parts of the machines should be replaced, etc.).
 b. *Dee Ø/IS knowing how to fix washing machines.

The example in (11a) provides further support that *be_{asp} V-ing* and *be_{aux} V-ing* (progressive) are different and that difference may be linked to *-ing* in both forms. The descriptive generalization is that *be_{asp}* is analyzed as introducing a habitual operator (HAB) into the logical representations. HAB binds variables ranging over eventualities. It relates an eventuality expressed by a predicate to an occasion. Consider the representation below:

- (12) a. Dee be riding her bike when the phone ring.
 b. HAB_e [ring (phone, e)] [riding bike (Dee, e)]

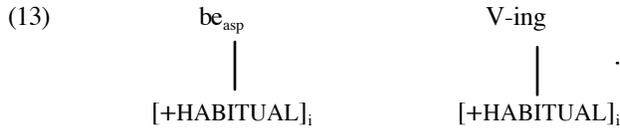
Informally, the representation in (12) says that habitually when the phone rings, Dee rides her bike at that time.

5 *-ing* as an Agreement Marker

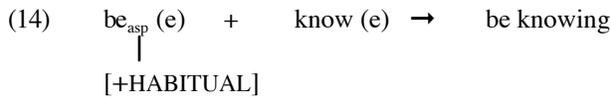
5.1 *-ing* and *Be_{asp}* Agreement

In the preceding section, it has been shown that *be_{asp} V-ing* and *be_{aux} V-ing* differ in several ways, and the differences raise the question about the role of *-ing* in the constructions. *Be_{asp} V-ing* can have the in-progress reading, but even when it has that reading, it still has the habitual or quantificational reading. In this construction, I want to suggest that *-ing* matches the habituality of *be_{asp}*, so it is analyzed as an agreement marker rather than as a marker of an event that is in progress. As an agreement marker, *-ing* is required in *be_{asp} V-ing* constructions; that is, *be_{asp}* forces the verb to occur with morphological agreement expressed as *-ing*:

⁵As is clear from the sentence *Dee is looking more and more like her mother these days*, stative verbs can occur in the progressive with a type of event reading.



The analysis of *-ing* as an agreement marker in be_{asp} *V-ing* constructions captures the difference in meaning between be_{asp} *V-ing* and be_{aux} *V-ing* constructions. Another advantage of such an analysis is that it provides an indirect explanation for why be_{asp} can never occur as \emptyset , but be_{aux} can. The explanation is that be_{asp} introduces into the representation an eventuality argument, so it cannot be absent. Habituality and the eventuality argument are linked to *-ing*:



On the other hand, in the case of be_{aux} *V-ing*, be_{aux} does not introduce any additional information into the structure, so it is not required to occur on the surface. Furthermore, if it is indeed the case that *-ing* in progressive constructions stativizes verbs, then it is clear why be_{aux} *-ing* is not compatible with stative verbs. On the other hand, *-ing* in aspectual *be* constructions coerces verbs to have a habitual reading, and this coercion applies to stative verbs by making them take on eventive readings.

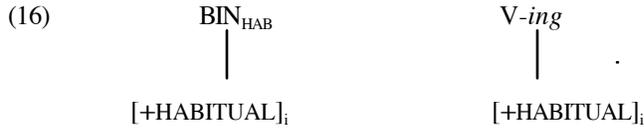
- (15) a. *Bruce is/ \emptyset knowing the answer.
 b. Bruce be knowing the answer.
 Literally: Bruce generally does something to show that he knows the answer.

Along these lines, be_{asp} must occur on the surface, not as a stativizer but as a quantificational element that introduces an eventuality argument into the representation and forces verbs to take a habitual interpretation. This description is compatible with the representation in (12b) above.

5.2 Extending the Agreement Analysis to BIN_{HAB}

The BIN_{HAB} *V-ing* reading is quite similar to the be_{asp} *V-ing* reading. Both constructions indicate habits, but the difference is that the former refers to habits that started in the distant past, while the latter does not make any claims about how far in the past the habit started. Given the similarity between the two constructions, it seems possible to extend the *-ing* analysis

proposed for be_{asp} V-*ing* constructions to *-ing* in BIN_{HAB} constructions. That is *-ing* in BIN_{HAB} constructions also matches the [HABITUAL] feature associated with BIN , as shown in the representation below:



A question that arises in light of the proposal for V-*ing* in BIN_{HAB} constructions is about the source of the [HABITUAL] feature. There are a number of explanations for this feature. The explanation that I will present is related to the cooccurrence restriction on be_{asp} and BIN . The claim is that both markers are generated in the head of the Aspect Phrase (AspP), so they cannot occur at the same time. The following are ruled out just for that reason:

- (17) a. *_[AspP] be BIN [_{VP} running] (*Bruce be BIN running.)
 b. *_[AspP] BIN be [_{VP} running] (*Bruce BIN be running.)

While the cooccurrence of be and BIN leads to ungrammaticality, the logical interpretation resulting from the sequence of the two is compositional and the actual meaning of habitual remote past, the reading of BIN_{HAB} . The meaning cannot be a result of the combination of the two markers because they never cooccur, but it might result from a habitual feature that may be left in the representation because there is no place for be_{asp} when BIN already occupies the Asp position. Because BIN also carries a pitch accent, then it is the marker that always wins out when there is competition between be_{asp} and BIN for the Asp position. Therefore, the sentences in (17) are ungrammatical. The meaning is conveyed by *Bruce BIN running*, and the argument here is that the feature [HABITUAL] occurs because the marker be_{asp} cannot. In this case, the *-ing* in BIN_{HAB} V-*ing* constructions is required by [HABITUAL] much like it is required in be_{asp} V-*ing* constructions.

The properties of progressive *-ing* in AAE that occurs in be_{aux} V-*ing* constructions is also found in BIN_{STAT} V-*ing* constructions (4a, b). The *-ing* in those constructions indicates that the eventuality indicated by the verb has been in progress from some point in the remote past to the speech time. In such cases, *-ing* can be argued to be a stativizer, which is different from the *-ing* that is required to match the [HABITUAL] that is associated with be_{asp} and BIN_{HAB} .

In summary, the argument is that *-ing* in be_{asp} and BIN_{HAB} V-*ing* constructions is a type of morphological agreement that is required by the fea-

ture [HABITUAL]. It does not have the same function as progressive *-ing* in *be_{aux}* and *BIN_{STAT}* V-*ing* constructions.

6 *-ing* Agreement and Acquisition of Aspectual Markers

Given the prominence of aspectual markers and the interpretation of verbal predicates in aspectual marker sequences, a number of questions about the development of these markers and their use with certain predicates in child AAE arise. In order to address some of these questions, data from two comprehension tasks, one on *BIN* and the other on *be_{asp}*, are considered.

6.1 *BIN* Comprehension Tasks

The *BIN* data are from forty-two three- to five-year-old developing AAE-speaking children in a child development program in southwest Louisiana.⁶ The participants were tested on ten scenarios that portrayed objects/characters as having been engaged in an activity for a long time as compared with other objects/characters that had been in the state or engaged in an activity for a shorter time. The scenarios consisted of short stories, pictures, and prompts. The interviewer read the short story to the participant while pointing to the corresponding pictures and asked *BIN* prompts (or target questions) related to the pictures. A sample scenario is given below:

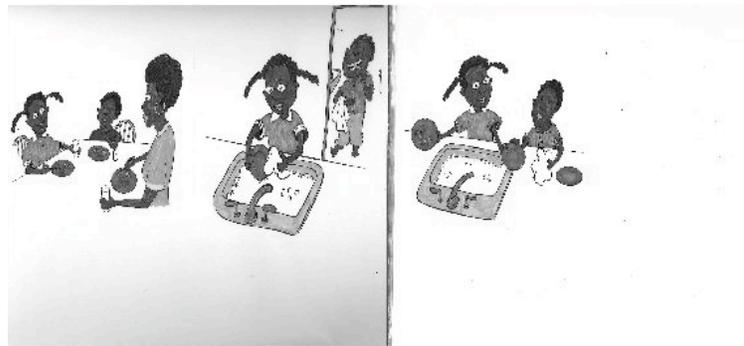


Figure 1 BIN Scenario (*BIN_{STAT}*)

Bruce and Jenny's mother told them that they could watch TV if they wash and dry the dishes after dinner. Jenny started washing the dishes while Bruce went to put on his pajamas and brush his teeth. Then he came back to help Jenny dry the dishes.

⁶Eight participants have been added to the study since the presentation. The overall results have not changed.

The ten scenarios consisted of *BIN* followed by *V-ing*, *V-ed*, *N*, and *Adj*; however, only *BIN V-ing* results will be reported here. The percentage correct for the *BIN V-ing* scenarios is given in the table below:

Prompt	% Correct
Scenario 1: Who <i>BIN</i> working at the kitchen sink? (Jenny) 'Who has been working at the kitchen sink for a long time?' (gloss)	83
Scenario 2: Who <i>BIN</i> knowing how to climb trees? (Jenny) 'Who has known for a long time how to climb trees?' (gloss)	81
Scenario 3: Who <i>BIN</i> fixing bikes? (the old man) 'Who has been fixing bikes for a long time?' (gloss)	55

Table 2 *BIN V-ing* Results

The participants scored over 80% correct on the *BIN working* and *BIN knowing* scenarios. The claim is that children who interpret *BIN* as a distant past marker, distinct from the simple past, will select Jenny for the *BIN working* scenario, and they will select the character who has known how to climb trees longer than the other characters for the *BIN knowing* scenario. In both cases, we get the *BIN_{STAT}* reading, in which a state has held from the distant past to the speech time. For example, according to the scenario, Jenny started working at the kitchen sink a long time ago, and the working at the kitchen sink event has held since that time. The participants scored much lower on the *BIN fixing* scenario. The difference between the *BIN fixing* scenario and the other *BIN V-ing* scenarios is that it is ambiguous between a *BIN_{STAT}* and a *BIN_{HAB}* reading. That is, in the *BIN fixing* scenario, the targeted character (old man) is not portrayed as fixing a bike at the speech time, so the story and pictures are compatible with the reading in which the old man started fixing bikes a long time ago, and he still fixes them from time to time. Compare the *BIN working* scenario, in which Jenny is portrayed as working at the kitchen sink, to the *BIN fixing* scenario, in which the old man is not working on the bike. If it is the case that children are sensitive to the *BIN_{STAT}* and *BIN_{HAB}* readings, then a possible explanation for the lower *BIN fixing* score is that the children have some difficulty with the *BIN_{HAB}* reading, in which *V-ing* is taken as a habitual agreement marker rather than as an in progress marker. The data for the *be_{asp}* scenarios offer some support for the hypothesis that children have difficulty with the habitual interpretation.

6.2 *Be_{asp}* Comprehension Tasks

The *be_{asp}* comprehension tasks were designed to determine whether developing AAE-speaking children associate *be_{asp}* with habitual situations. The *be_{asp}* data are from twenty-five developing AAE-speaking children in a child development program in southwest Louisiana.⁷ The six test scenarios included a short story, corresponding pictures, and a prompt or target question that featured *be_{asp}* followed by a verbal or non-verbal predicate. A sample scenario is given below:



Figure 2 *Be_{asp}* Scenario

At lunchtime, all the kids eat together. Bruce always has turkey sandwiches because he loves turkey. He had turkey sandwiches last week and this week. Jenny likes peanut butter and jelly or ham and cheese. She doesn't eat turkey for lunch. Faye likes everything. She sometimes has a cheese sandwich. Today, Faye has a turkey sandwich but Bruce doesn't. He has soup. *Who be having turkey sandwiches for lunch?*

As the results show, the participants scored much higher on the *BIN* scenarios than they did on the *be_{asp}* scenarios:

⁷Nine participants have been added to the study since the presentation.

Prompt	% Correct
Scenario 1: Who be swimming in the neighborhood pool? 'Who usually swims in the neighborhood pool?' (gloss)	48
Scenario 2: Who be having turkey sandwiches for lunch? (Bruce) 'Who usually has turkey sandwiches for lunch?' (gloss)	56
Scenario 3: How does Faye be getting to school? 'How does Faye usually get to school?' (gloss)	56
Scenario 4: Where does Jenny's sister Haley be hiding? 'Where does Jenny's sister Haley be hiding?' (gloss)	60

Table 3 *Be_{asp}* V-*ing* Results

Given the function of *be_{asp}*, there is no ambiguity; all of the constructions have a habitual reading. While the participants scored above chance on all of the scenarios, they did not score above 60%. The result for *BIN fixing* (Table 2) is closer to the result for the *be_{asp}* scenarios than it is to the results for the *BIN_{STAT}* V-*ing* scenarios. If the child speakers have not quite grasped the feature [HABITUAL], then they would be expected to do less well on *BIN_{HAB}* and *be_{asp}* scenarios than on the other scenarios. These are the findings; however, because there is only one *BIN_{HAB}* scenario in the *BIN* experiment, it is not clear how reliable the claim about *BIN_{HAB}* is. However, the results for *be_{asp}* seem to be more robust. Nevertheless, given the claim about [HABITUAL] requiring an *-ing* agreement marker, more data and experimentation on these constructions would be useful in providing insight into children's interpretation of the habitual markers and the type of morphology that accompanies them.

7 Summary

V-*ing* occurs in progressive contexts as well as in *be_{asp}* contexts, and there is sufficient evidence to show that *-ing* in the two contexts has different functions.

-ing in *be_{aux}* V-*ing* constructions indicates that an eventuality is in progress, while *-ing* in *be_{asp}* V-*ing* contexts is a type of agreement required by the habitual marker *be_{asp}*. In other words, in the latter context, *-ing* is a morphological marker that agrees with the [HABITUAL] feature of *be_{asp}*, which can coerce even stative predicates into a habitual reading. The question about whether *-ing* in *BIN_{HAB}* V-*ing* constructions is also a type of

morphological agreement that is required by the [HABITUAL] feature was also raised. Data from developing AAE-speaking children show that they fare well on BIN_{STAT} *V-ing* constructions; however, they do less well on be_{asp} *V-ing* and BIN_{HAB} *V-ing*. The claim is that children have difficulty with the habitual feature associated with the aspectual markers, but more experimental work should be conducted to determine whether children encounter the same type of problems with be_{asp} and BIN_{HAB} .

References

- Baugh, J. 1980. A Reexamination of the Black English Copula. *Locating Language in Time and Space*, ed. W. Labov, 83-106, New York: Academic Press.
- Déchainé, R-M. 1993. Predicates Across Categories. Doctoral dissertation. University of Massachusetts.
- Green, Lisa. 1993. Topics in African American English: The Verb System Analysis. Doctoral dissertation. University of Massachusetts.
- Green, L. 1998. Remote Past and States in African-American Vernacular English. *American Speech* 73:115-138.
- Green, L. and T. Roeper. To appear. The Acquisition Path for Tense-Aspect: Remote Past and Habitual in Child African American English. *Language Acquisition*.
- Labov, W. 1969. Contraction, Deletion, and Inherent Variability of the English Copula. *Language* 45; 715-762.
- Parson, T. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge: MIT.
- Rickford, J. R., A. Ball, R. Blake, R. Jackson, N. Martin, 1991. Rappin' on the Copula Coffin: Theoretical and Methodological Issues in the Analysis of Copula variation in African-American Vernacular English. *Language Variation and Change* 3, 102-132.
- Schubert, L. K. and F. J. Pelletier. 1989. Generally Speaking, or Using Discourse Representation Theory to Interpret Generics. *Properties, Types and Meaning*, Vol. 2: Semantic Issues, ed. G. Chierchia, B. H. Partee, and R. Turner. 193-268. Dordrecht: Kluwer.
- Smith, C. S. 1997. *The Parameter of Aspect*. (2nd edn). Dordrecht: Kluwer Academic Press.
- Walker, J. A. 2000. Rephrasing the Copula: Contraction and Zero in Early African American English, *The English History of African American English*. ed. S. Poplack. 35-66. Oxford: Blackwell.

Steal Me an Apple: Version in Georgian

OLGA GUREVICH

This paper examines a set of valency-affecting constructions in Georgian marked by pre-stem vowels on finite verb forms, known as *version*. I explore the morphosyntactic and semantic functions of each version construction and the relationship between these constructions and the larger morphosyntactic context.¹

1. Introduction

This paper examines a set of valency-affecting constructions in Georgian marked by pre-stem vowels on finite verb forms. These constructions, known collectively as *version*², are at the core of Georgian morphosyntax. Aside from the common morphological distribution and general relation to valency, version constructions can be quite diverse. I explore the morphosyntactic and semantic functions of each version construction and how these functions are determined and affected by the larger syntactic and semantic contexts in which they are embedded.

Version vowels are used for a variety of valence-related functions, which are the focus of this paper. Examples include valency-increasing marking of beneficiaries, reflexivity, marking of passives/impersonals,

¹I wish to thank Jim Blevins, Nino Amiridze, and Kevin Tuite for helpful suggestions on earlier drafts of this paper. Thanks to the audience at TLS 9, and particularly Rose-Marie Déchaine, for drawing my attention to some cross-linguistic correlates of version. Finally, I am grateful to Shorena Kurtsikidze, Vakhtang Chikovani, and Vano Nasidze for providing much of the Georgian data. All errors are my responsibility.

²The term ‘version’ is a translation of the Georgian *kceva* meaning ‘change’.

causatives, and the marking of indirect object agreement. In addition, version vowels have in some cases become grammaticalized and perform purely morphological functions, although such cases will not be discussed here in detail.

1.1 Version as Participant Affectedness

The data described in this paper suggest that for most cases, the slot in which version vowels appear has a coherent general function. Overall, the elements in this slot indicate changes in valence and grammatical voice. However, the syntactic arguments over which such elements take scope, and their precise function, are determined by the larger syntactic and semantic context in which the verb forms appear, as well as the lexical semantics of the verbs. The functions of the version elements can be further subdivided into several subsystems.

The most prototypical and most often described uses of version vowels involve the marking of *participant affectedness*, meaning that some discourse participant is directly affected by the action denoted by the verb. This term is related to the cognitive notion of *salience* or *prominence* (cf. Langacker 1991).

Version as participant affectedness is most often seen on active verbs, where version vowels differentiate the largest number of types of affectedness. Here, objective version (affected participant is the indirect object), subjective version (subject), and locative version (indirect object) are possible. Version markers can also express participant affectedness on passive and impersonal verbs, but, as will be demonstrated below, the types of affectedness distinguished for these verbs are more restricted and less differentiated than for active verbs. In each of these cases, the function of version vowels crucially depends on the surrounding morphosyntactic contexts.

The body of the paper describes the various functions of version vowels, following a brief overview of Georgian verb morphosyntax.

2. Morphosyntactic preliminaries

The properties of Georgian morphosyntax relevant to understanding version include verb conjugation classes, the tense/aspect/mood (TAM) system, and the morphotactic structure of the Georgian verb.

2.1 Conjugation Classes

Georgian verbs are traditionally divided into conjugation classes. Following Harris (1981) and Aronson (1990), I distinguish four conjugation classes identified by morphological criteria. The classes largely correspond to valence generalizations, though there are notable exceptions.

SERIES I		SERIES II (AORIST)	SERIES III (PERFECT)
PRES. SUBSERIES	FUT. SUBSERIES		
Present	Future	Aorist	Perfect
Imperfect	Conditional		Pluperfect
Present subjunctive	Future subjunctive	Aorist subjunctive	(Perfect subjunctive)

TABLE 1 Series and screeves (“Present”, “Future”, etc. are screeves)

The conjugation classes are as follows.

- Conjugation I: “Transitive”; mostly active transitive verbs, e.g. ‘build’, ‘draw’.
- Conjugation II: “Unaccusative (Passive)”; mostly intransitive, non-active verbs often derived from transitives, e.g. ‘be written’. Includes passive and impersonal verbs.
- Conjugation III: “Unergative (Medioactive)”; mostly intransitive, active verbs, e.g. ‘dance’, ‘walk’.
- Conjugation IV: “Indirect”; mostly transitive, non-active verbs where the subject is the experiencer of the action, e.g. verbs of experience and possession, e.g. ‘like’, ‘have’. The grammatical subject of these verbs is marked by dative case and by object markers on the verbs.

2.2 TAM System

Georgian verbs inflect according to a set of tense, aspect, and mood (TAM) parameters. A collection of TAM properties is called a *screeve*, from *mc’k’rivi* ‘row’; it is similar to a traditional inflectional paradigm with a collection of cells for different combinations of subject and object properties with the same TAM values.

A collection of screeves with common morphosyntactic properties is called a *series*. There are three distinct series in Georgian. Table 1 shows the distribution of screeves in series.

Series III of transitive and unergative, and all series of indirect verbs have dative-marked subjects and nominative-marked objects. This phenomenon is known as *inversion*, and its relation to version vowels is discussed in section 9..

2.3 Structure of the Georgian Verb

An inflected Georgian verb has a fairly complex structure with up to eleven distinct slots. The structure of the verb is most commonly described in terms of a template, as in (1).

- (1) Simplified verb template in linear order:

(PV₁)-(Pron1₂)-(Ver₃)-root₄-(TS₅)-(Scr₆)-(Pron2₇)³

The functions of the slots are briefly explained in (2).

(2) Functions of other slots:

1. Preverbs (PV), e.g. *uk'u-* 'back', *she-* 'into', *mo-* 'towards the speaker', indicate aspect and spatial orientation.
2. Prefixal pronominal marker (PRON1), e.g. *v-* '1SUBJ', *m-* '1SGOBJ'
3. Version vowel (VER), e.g. *i-*, *u-*, *a-*, *e-*
4. Root, e.g. *xat'* 'draw'
5. Thematic suffix (TS), e.g. *-av*, *-eb*, has no independent meaning and indicates lexical class.
6. Screeve (tense) marker (SCR), e.g. *-i*, indicates inflectional class and TAM category.
7. Suffixal pronominal marker (PRON2), e.g. *-s* '3SGSUBJ', *-en* '3PLSUBJ', *-t* 'PL'

Not every verb has all of these parts, and the only required part is the root.

The version vowels *i*, *u*, *e*, *a* occupy a slot immediately before the root. Only one version vowel can occupy the slot. The function of the version vowels depends on the larger syntactic context in which it is used, i.e. verb conjugation class, valency, and TAM.

The types of version constructions are described below. They have been grouped according to the type of function served and the type of morphosyntactic context in which these uses occur.

3. Active Verbs—Objective Version

'Objective version' indicates that the event is directed towards and/or affects a participant which is most often coded as an indirect object. Objective version is the least lexically restricted type of version, and is most often optional. It is attested on transitive (Conjugation 1) verbs, and some active intransitive (Conjugation 3) verbs.

3.1 Objective Version in Transitive Verbs

If the indirect object is 1st or 2nd person, the version vowel is *i*. In (3), the beneficiary of the action (1sg pronoun) is expressed in a postpositional phrase. By contrast, in (3b), the beneficiary is encoded as

³Abbreviations: Aor=aorist; Dat=dative; Erg=ergative; Nom=nominative; Obj=object; Pass=passive; Pl=plural; Pron=pronominal marker; PV=preverb; Scr=screeve marker; Sg=singular; Subj=subject; TS=thematic suffix; Ver=version

an indirect object; the verb agrees with it by means of the pronominal affix *m-*, and indicates the affectedness relationship with the version vowel *i-*.

- (3) a. meri-m da-xat'a **chem-tvis** surat'i
 Mary-ERG PV-paint.AOR **1sg-for** picture-NOM
 'Mary painted a picture for me.'⁴
- b. meri-m da-**m-i**-xat'a (**me**) surat'i
 Mary-ERG PV-1SGOBJ-**ver**-paint (**1sg**) picture-NOM
 'Mary painted a picture for me.'

If the indirect object is 3rd person, the version vowel is *u-*, as in (4). There is no indirect object marker on the verb. It has been suggested that *u* is a fusion of 3rd person object marker *s/h-* and *i-*, but there is no independent phonological evidence for this.

- (4) meri-m da-**u**-xat'a (**mas**) surat'i
 Mary-ERG PV-**ver**-paint (**3sg**) picture-NOM
 'Mary painted a picture for him/her.'

The syntactic effect of objective version is to elevate the affected participant to 'core argument' status. It is marked by the dative case and is cross-referenced on the verb by object markers. The direct object is often possessed by the indirect object, or becomes his/her possession. The indirect object is usually animate. The thematic role of the indirect object is often a recipient or beneficiary.

The discourse factors affecting the use of objective version are discussed in more detail in section 3.4.

3.2 Objective Version in Ditransitive Verbs

When objective version is used with a ditransitive verb, the syntactic effect is the addition of a (second) indirect object. The object markers on the verb co-reference the same direct object referred to by version. In (5b) and (5c), the use of a version marker coincides with the addition of a fourth argument to the verb, essentially creating two indirect objects. Pronouns are often dropped in Georgian, but the argument structure can be recovered from verb agreement.

- (5) a. man mo-**h**-p'ara (mas) vashl-i
 he.ERG PV-**1sgObj**-steal.AOR (3SG.DAT) apple-NOM
 'He stole an apple **from him/her**.' (Boeder 1969)

⁴Unless otherwise indicated, all data are from native-speaker elicitation by the author.

- b. man mo-**m-i-p**'ara mas vashl-i
 he.ERG PV-**1sgObj-ver-steal**.AOR 3SG.DAT apple-NOM
 'He stole an apple **for me** from him/her.'
- c. gtxov, (me) gada-**m-i-c-e** es c'igni
 please, (1SG) PV-**1sgObj-ver-give**-AOR this book.NOM
 P'avles!
 Paul.DAT
 'Please give this book to Paul for me!' (Boeder 1969)

3.3 Objective Version in Intransitive Verbs

The uses of objective version with intransitive verbs are more restricted than with transitive verbs. The result of the action is always directed at the affected participant, but exactly how this is done is determined by the lexical semantics of the verb, and not all verbs can combine with objective version.

In many cases, the impossibility of a version marker seems to be motivated by the lack of a semantic connection between the affected participant and the object / result of the verb's action.

In (6a), the objective version form of the verb means to perform the dancing action for someone; this is the most productive use of objective version consistent with transitive examples above.

- (6) a. shen m-**i-cek**'v-eb (me)
 you 1SG**OBJ-ver-dance**-TS (1SG)
 'You will dance **for me**.'
- b. shen **u-cek**'v-eb (mas)
 you **ver-dance**-TS (3SG.DAT)
 'You will dance **for him**.'

In (7b), the objective version verb form becomes transitive and idiomatic, meaning to make someone fly / go away.

- (7) a. da-v-pren
 PV-1SUBJ-fly
 'I will fly.'
- b. exla da-g-i-pren shen
 now PV-2OBJ-VER-fly 1SG
 'I will fly you away (=make you go away).'

In other cases, the only context in which intransitive Conjugation 1 verbs allow objective version is by adding an object which is somehow related to/possessed by the affected participant (8b), thus making the verb transitive.

- (8) a. *shen me m-i-cxovr-eb
 2SG 1SG 1SGOBJ-VER-live-TS
 ‘You (will) live for me.’
- b. am bich’s tu k’argad m-i-cxovr-eb, me
 this boy.DAT if well 1SGOBJ-VER-live-TS, 1SG
 da-g-a-sachukreb
 PV-2OBJ-VER-reward-TS
 ‘If you give this boy a good life for me, I will reward you.’

Finally, in some cases the use of version forms implies a possessed object even when it is not overtly present (9).

- (9) mo-m-i-mat’-eb
 PV-1OBJ-VER-increase-TS
 ‘you will increase for me (e.g., my salary)’

3.4 Discourse Factors Relevant for Objective Version

The indirect object is usually affected via some connection to the object or the result of the action. The object is very often inalienably possessed by the affected participant, as in (10).

- (10) is g-i-t’ex-s shen mk’lavs
 he 2OBJ-VER-break-3SG your arm.DAT
 ‘He breaks (your) arm for you.’

The object can be in an animate relation to the affected participant like relatives or children (11, 12)

- (11) she-m-i-k-e me shvili
 PV-1SGOBJ-VER-praise-AOR 1SG child.NOM
 ‘You praised my child for me.’
- (12) is kali ga-m-i-gizh-eb-s bavshvebs
 this woman.NOM PV-1SGOBJ-VER-crazy-TS-3SG child.PL.DAT
 ‘This woman will drive (for me) my kids crazy.’

Finally, the object can simply be in possession of the affected participant. In such cases, the version form implies possession (13a); the *-tvis* form implies non-possession (13b). A non-version construction, if possible, has a different meaning from its version counterpart, and the meaning difference is motivated by the general meaning of “participant affectedness” carried by version.

- (13) a. is g-i-t'ex-s shen doks
 he 2OBJ-VER-break-3SG your clay-jug.DAT
 'He breaks a clay jug for you.' (Implication: your jug)
- b. is shentvis doks t'ex-s
 he you.for clay-jug.DAT break-3SG
 'He breaks a clay jug for you.' (Implication: not your jug)

When there is a choice between using a version and a non-version construction, the general deciding factor seems to be the degree to which a participant is affected by the action. Greater degree of affectedness usually corresponds to use of version, whereas lesser degree of affectedness corresponds to a *-tvis* postpositional phrase or some other device. This choice works both ways: given that a version construction was used, the interpretation has to be that of more affectedness. In addition, the non-version constructions can sometimes be interpreted as either the more affected or the less affected option, whereas the use of version unambiguously signals the more affected variant.

The more specific parameters affecting the choice are as follows, with the more affected variant on the right:

- Possession by indirect object vs. non-possession
- Inalienable possession vs. alienable possession
- Goal / recipient of action vs. beneficiary of action

In some cases, the postpositional and version constructions are equivalent and freely interchangeable. However, the version construction is preferred when the affected participant is directly linked to the action denoted by the verb, directly affected by it, and/or present at the event. The interpretation of (14a) is ambiguous between a recipient (more affected) and a beneficiary; however, the recipient interpretation sounds odd in (14b) and is impossible in (15b).

- (14) a. shen **m-i**-mgheri.
 2SG **1sg-ver**-sing.PRES.2SG
 'You sing **for me / to me.**'
- b. shen **chem-tvis** mgheri.
 2SG **1sg-for** sing.PRES.2SG
 'You sing **for me / ?to me.**'

The difference in affectedness is stronger for verbs that may include an implicit affected participant.

- (15) a. is **me m-i**-q'viris.
 3SG **1sg 1sg-ver**-yell.PRES.3SG

meaning of the verb form is non-compositional. Rather, the presence of the vowel *i-* in an active verb form in conjunction with an indirect object and corresponding verb markers can be interpreted as objective version. On the other hand, the presence of *i-* in conjunction with a subject marker and no affected indirect object marks subjective version. In other words, the larger construction and the combination of other morphs determines the function of the version vowel.

5. Active Verbs—Locative Version

Locative version (also known as *superessive*, cf. Hewitt 1995), indicates that the action was performed in some spatial relation to the indirect object. The spatial relation is usually ‘onto’, but metaphoric extensions are also possible. Locative version is marked by version vowel *a-*.

In (24a), the landmark of the breaking action (his head) is encoded as an indirect object, and the verb has the version vowel *a-*. The same notion can be paraphrased with a postpositional phrase as in (24b).

- (24) a. me v-**a**-t'ex-av j'ox-s mis tav-s
 I 1SUBJ-**ver**-break-TS stick-DAT his head-DAT
 ‘I break a stick **over** his head.’
- b. me v-t'ex-av j'ox-s mis tav-**ze**
 I 1SUBJ-break-TS stick-DAT his head-**on**
 ‘I break a stick **over** his head.’

The spatial relation can be interpreted metaphorically, as in (25).

- (25) (me) (shen) mo-g-**a**-mghere
 (1SG) (2SG) PV-2OBJ-**ver**-sing.AOR
 ‘I sang you off / sang as you were leaving.’

Locative version appears to be restricted to transitive verbs, but it is too rare for any definitive conclusions.

6. Version as Participant Affectedness, Non-Active Verbs

Version can also indicate participant affectedness and valency changes in some passive and impersonal verbs in Conjugation 2. However, the types of distinctions marked by version in Conjugation 2 are different from those marked in Conjugation 1. In particular, the types of affectedness are not differentiated to the same extent as in Conjugation 1. The main point of this section is to demonstrate the difference between version in the two conjugation classes, and to argue that the larger

contexts (conjugation class and series) again determine the function of version vowels.

Conjugation 2 includes many passive or impersonal counterparts of Conjugation 1 verbs. There are three ways to form Conjugation 2 verbs: by addition of a pre-stem vowel, addition of suffix *-d*, and without adding any extra elements. The two latter verb types follow the same pattern as active verbs with regards to version vowels.

For passives with prefixal formation, version vowels *-i-* and *-e-* can mark the presence (*-e-*, 26b) or the absence (*-i-*, 26a) of an affected participant, encoded as an indirect object.

- (26) a. *i-xat'eba*
 ver-draw.INTR.PRES
 ‘It is being drawn.’
- b. *e-xat'eba* (mas)
 ver-draw.INTR.PRES (he.DAT)
 ‘It is being drawn **for / in front of / on him/her.**’

The syntactic function of *e-* is to add an indirect object to the verb frame. Only the presence of an affected participant is marked; the way in which this participant is affected is either not specified or is provided by the lexical semantics of the verb. Further, subjective and locative versions are not possible. In a sense, the version vowel *e-* combines the functions of objective and locative versions.

Example (27) demonstrates that Conjugation 2 verbs are not always passive. Rather, this conjugation signals that the subject is not entirely in control, or the direct object is not prominent.

- (27) *i-landzgheba*
 VER-curse.PASS
 ‘(S)he is engaged in cursing’ (not relevant who (s)he is cursing).
 (Nino Amiridze, p.c.)

Conjugation 2 verbs often become lexicalized as allowing only one or the other version marker. In (28b), the root meaning ‘to use’ plus version marker *-e-* have become lexicalized as a bivalent verb ‘to help’. This verb is no longer transparently connected to the verb ‘use’, and thus removing the version marker would take away its core meaning.

- (28) a. *is shen g-e-xmar-eba*
 3SG 2SG 2OBJ-VER-help-SCR
 ‘He helps you(sg).’

- b. is i-xmar-eb-a
 3SG VER-help-SCR
 ‘It is being used / (*help).’

When both vowels are possible, *-e-* carries a very strong implication of an added affected participant, even when that participant is not overtly present in the sentence as an indirect object. In (29b), the implied second participant (1sg pronoun) is not present in the sentence. It does not trigger verbal agreement and therefore cannot simply be the result of pro-drop.

- (29) a. she-i-k’vreb-i
 PV-VER-group-SCR
 ‘You will group yourself / concentrate.’
 b. she-e-k’vr-ebi
 PV-VER-group-SCR
 ‘You and I will group / do something “under the table” together.’

The affected participant interpretation is often blocked by more strongly lexicalized inceptive or patientive interpretations. In (30), the addition of the version marker *-e-* is interpreted according to the productive pattern, meaning that the action affects some participant X. In (30), there is a strongly lexicalized interpretation of ‘feeling like X’ and something appearing to someone.

- (30) a. i-xat’-eba
 VER-draw-SCR
 ‘It is drawn.’
 b. e-xat’-eba
 VER-draw-SCR
 ‘It is drawn for him/her.’
 c. m-e-xat’-eba
 1SGOBJ-VER-draw-SCR
 ‘(*It is drawn for me) / I feel like drawing / It (an image) appears to me.’

In (31b) and (31c), the lexicalized meaning of ‘register’ is more entrenched, and the simple beneficiary interpretation of the form with *-e-* is completely impossible.

- (31) a. i-c’er-eba
 VER-draw-SCR
 ‘It is written.’
 b. e-c’er-eba
 VER-write-SCR

‘(*It is written for him/her) / (S)he is registering / feels like writing / It will be written.’

c. m-e-c’er-eba

1SGOBJ-VER-write-SCR

‘(*It is written for me) / I am registering / I feel like writing.’

Many Conjugation 2 verbs share roots with Conjugation 1 verbs. However, it is unclear how they came to have the vowels *i-* and *e-*. It is plausible that the use of *i-* is related to subjective version (reflexive) *i-* in Conjugation 1 verbs: it is quite common cross-linguistically for passive, impersonal, and reflexive verbs to have the same morphology (cf. Klaiman 1991, Blevins 2003). The use of the same morphological marker for directedness (as in objective version) and lack of volitionality (as in Conjugation 2 verbs) is also not unprecedented: a somewhat similar phenomenon is described for the Salish language family (Demirdache 1997). However, the direction of the historical development in Georgian is not known.

7. Causativization

The version vowel *-a-* can be used in forming causatives, accompanied by the thematic suffix *-eb*. Causatives formed from intransitive verbs are illustrated in (32).

(32) Intransitive causatives:

a. v-q’ep

1SUBJ-bark

‘I bark.’

b. v-a-q’ep-eb

1SUBJ-**ver**-bark-**ts**

‘I make him bark.’

Causatives formed from transitive verbs may contain an extra suffix *-in*, as in (33b).

(33) Transitive causatives:

a. v-c’er mas

1SUBJ-write it.DAT

‘I write it.’

b. v-a-c’er-**in-eb** mas

1SUBJ-**ver**-write-**caus-ts** it.DAT

‘I make him write it.’

The same pattern, including the version vowel *a-* and the thematic suffix *-eb*, is also a very common way of forming verbs from adjective and noun stems. This class contains around 2700 verbs and is quite

productive (Melikishvili 2001). Historically, the use of *a-* in these cases may signal a verbalizing function (an increase in valency as compared to a noun or adjective). However, many of the verbs in this class no longer share a synchronic connection with the stem from which they were derived.

In this class of verbs, the version vowel *a-* often occurs in complementary distribution with other version vowels. Most verbs in this category belong to Conjugation I, as in (34a). When the pre-stem slot is occupied by version vowel *a-*, it cannot express version-like meanings, so that beneficiaries of actions have to be expressed in postpositional phrases (34b). However *a-* can be replaced by other version vowels to form objective or subjective version, as in (34c).

- (34) a. saxl-s **a**-shen-**eb**
house-DAT **ver**-build-**ts**
‘You build a house (beneficiary not specified).’
- b. saxl-s **a**-shen-**eb** mis-tvis / *mas
house-DAT **ver**-build-TS he.GEN-for / *he.DAT
‘You build a house for him.’
- c. saxl-s **u**-shen-**eb** *mis-tvis / mas
house-DAT **ver**-build-TS *he.GEN-for / he.DAT
‘You build a house for him.’

Because it stands in contrast with other possible version vowels, *a-* has been called ‘neutral version’ in the Georgian linguistic tradition (Boeder 1969, Hewitt 1995). However, its meaning and function are not related to the ‘participant affectedness’ meanings of the other version vowels, and it is not required on all verbs.

8. Version Markers as Experiencer Agreement

8.1 *i/u-*

For many verbs, version vowels have become grammaticalized as experiencer agreement. Such verbs subcategorize for an indirect object and may or may not have a direct object. This may be an intermediate stage between semantically / pragmatically motivated objective version and fully grammaticalized, morphological uses of version vowels.

The verbs in (35) do not subcategorize for a direct object and belong to Conjugation 3.

- (35) a. m-i-chivl-**eb**
1SGOBJ-VER-sue-TS

- ‘You will sue me.’
 b. m-i-chkmet’ xelze
 1SGOBJ-VER-pinch hand.on
 ‘You will pinch me on the hand.’

In (36), the verb ‘approve’ is phrasal, consisting of a required direct object, and the patient is encoded as an indirect object and cross-referenced on the verb with a version vowel.

- (36) k’rebam da-**u**-ch’ir-a mxari am
 council.ERG PV-**ver**-hold-SCR shoulder.NOM this
 k’andidat’uris sheq’vanas saarchevno biulet’enshi.
 candidacy.GEN inclusion.DAT electoral ballot.in
 ‘The council approved the inclusion of this candidacy in the
 electoral ballot.’ (Internet)

8.2 *s/h-*

Object agreement in Georgian also sometimes includes the prefix *s/h-*, where the two consonants are phonologically conditioned allomorphs. The prefix cross-references 3rd person indirect objects and cannot co-occur with a version vowel (37a); the indirect object can fulfill a number of thematic roles. Often, constructions which use *s/h-* can be contrasted with other constructions in which the affected participant is expressed in a postpositional phrase, similar to the contrast between version vowel-containing constructions and those with *-tvis* phrases (37b).

- (37) a. man mo-**h**-p’ara (mas) vashl-i
 he.ERG PV-**s/h**-steal.AOR (3SG.DAT) apple-NOM
 ‘He stole an apple from him/her.’
 b. man mo-p’ara mat-gan vashl-i
 he.ERG PV-steal.AOR 3SG.GEN-from apple-NOM
 ‘He stole an apple from him/her.’

Thus, in some ways, the constructions with *s/h-* can serve to elevate a participant to an indirect object (core argument) status, and are functionally similar to version vowels. Such classification has been suggested by Shanidze and by Boeder (p.c.). The forms with *s/h-* are not as common in modern Georgian as they once were, are less productive than forms with version vowels, and their lexical distribution requires a more thorough investigation. However, it does seem that such forms, along with *i/u-* object agreement forms, serve as the morphological basis for the inversion constructions discussed in the next section.

9. Inversion

The inversion construction de-emphasizes the agentivity of the subject. Inversion can occur in series III (perfect and pluperfect screeves) of active verbs (i.e. Conjugation 1 and 3) and in all forms of Conjugation 4 ('indirect') verbs. Inversion in series III coincides with an evidential interpretation of the perfect and pluperfect screeves. Indirect verbs do not have evidential semantics, but tend to include patient-like verbs, e.g. verbs of experience, perception, and possession.

In inversion, semantic subjects appear in the dative case and semantic direct objects appear in the nominative. Agreement markers of the verbs are correspondingly flipped. Harris (1981), working within the framework of Relational Grammar, has argued that the initial subject in inversion is realized as an indirect object (see Blevins (forthcoming) for a rendition of this analysis in Lexical Mapping Theory). It is a matter of some theoretical debate whether the syntactic phenomenon is the same in series III and in indirect verbs (Tuite, p.c.), but its effect on case-marking, version vowel use, and agreement is the same in both cases. Under any interpretation, dative subjects in inverted series III and indirect verbs have some properties in common with indirect subjects in non-inverted contexts, a fact that was also pointed out by Tschenkeli (1958) and Shanidze (1973).

9.1 Inversion in Active Verbs

Instead of normal version vowels, the pre-stem slot in these verbs has obligatory markers *i/u-* or *e-*, which indicate screeve (TAM) in series III. In the perfect, forms with a 1st and 2nd person initial subject have the version vowel *i-* (38a); forms with a 3rd person initial subject have the version vowel *u-* (38b). A beneficiary of the action can only be expressed in a postpositional phrase.

- (38) a. (turme) saxl-i a-m-i-sheneb-ia.
 (apparently) house.NOM PV-1SGOBJ-**ver**-build-PERF
 'I have (apparently) built a house (for a friend).'
- b. mas saxl-i a-u-sheneb-ia.
 3SG.DAT house.NOM PV-**ver**-build-PERF
 '(S)he has built a house (for a friend).'

In the pluperfect screeves, all forms have the version vowel *e-* (39).

- (39) a. saxl-i (amxanag-is-tvis) unda
 house.NOM (friend-GEN-for) must
 a-m-**e**-sheneb-ina.
 PV-1SGOBJ-**ver**-build-PLUPERF

- ‘I was supposed to have built a house (for a friend).’
- b. saxl-i (amxanag-is-tvis) unda a-e-sheneb-ina.
 house.NOM (friend-GEN-for) must PV-**ver**-build-PLUPERF
 ‘(S)he was supposed to have built a house (for a friend).’

Subjective or objective versions are not possible in inversion (40).

- (40) *saxl-i amxanag-s unda
 house.NOM friend-DAT must
 a-m-**i/u**-sheneb-ina.
 PV-1SGOBJ-**ver**-build-PLUPERF
 ‘I was supposed to have built a house for a friend.’

In inversion, the semantic subject appears to have some the formal properties of an indirect object, including version vowels. This connection may have arisen through the types of verbs described in section 8.. The version vowels thus have some connection to the syntax, and are not simply morphological or lexical. However, the syntactic function of the version vowels does not explain why *i/u-* is used in the perfect and *e-* is used in the pluperfect.

9.2 Inversion in Indirect Verbs

Some indirect verbs can have an obligatory marker *i/u-* or *e-* in series I and II (41). Their series III forms are usually participial and thus do not contain version vowels.

- (41) a. m-**i**-q'vars / **u**-q'vars
 1SGSUBJ-**ver**-love / **ver**-love
 ‘I love it / (S)he loves it.’
- b. m-**e**-mgher-eba
 1SGSUBJ-**ver**-sing-PASS
 ‘I feel like singing.’

Other indirect verbs use object markers and *s/h-*, although the future tends to be marked by *e-* for all indirect verbs (42).

- (42) a. mo-**m**-c'ons / mo-**s**-c'ons
 PV-1SG-like / PV-3SG-like
 ‘I like it / (S)he likes it.’
- b. mo-**e**-c'oneba
 PV-**VER**-like
 ‘I will like it.’

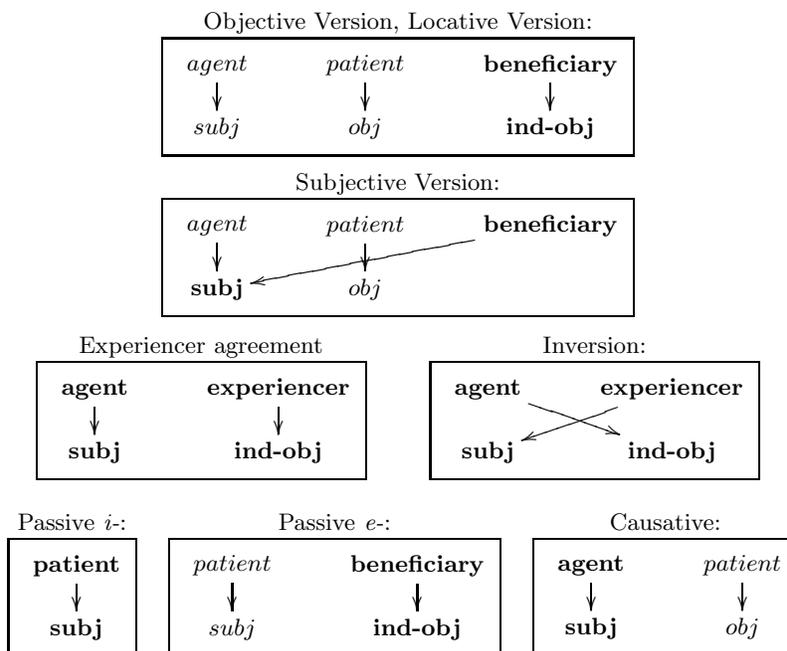


FIGURE 1 Argument-linking representation of version constructions

The contrast between *i/u*- and *s/h*- is reminiscent of the contrast between these markers in grammaticalized patient marking. In the case of indirect verbs, however, the motivation for using one construction over another seems to be even more lexicalized.

10. Conclusion

The pre-stem slot on the verb is most often related to valency and voice. Most of the syntactic differences between different version constructions can be represented in terms of linking between thematic roles and syntactic arguments, as in Figure 1. The parts of the linking structure that are contrasted with other version constructions or constructions without version vowels are in bold-face. Additional semantic differences, such as the distinction between objective, locative, and *s/h*-version are not indicated in the linking diagrams.

This generalization is not without exception. In some contexts, pre-stem vowels may have purely morphological functions, quite unrelated to syntactic valency or voice. The so-called ‘neutral’ version *a*- in section

7. is one such example; others abound, but their description is beyond the scope of this paper.

On the other hand, the pre-stem slot is not the only part of a verb form involved in expressing valency changes or voice. As mentioned in section 7., thematic suffixes (especially *-eb*), and the suffix *-in* also participate in the formation of causatives. Some passive and impersonal verbs, on the other hand, are formed by adding the suffix *d-* instead of a version vowel. In short, the valency-changing function is not uniquely associated with the pre-stem slot—there are exceptions on both formal and functional sides of the relationship.

References

- Aronson, Howard I. 1990. *Georgian. A Reading Grammar. Corrected Edition*. Columbus, OH: Slavica Publishers.
- Blevins, James P. 2003. Passives and impersonals. *Journal of Linguistics* 39:473–520.
- Blevins, James P. forthcoming. Thematic inversion in georgian.
- Boeder, Windried. 1969. Über die versionen des georgischen verbs. *Folia Linguistica* 2:82–152.
- Demirdache, Hamida. 1997. Out of control in st'at'imets. In A. Mendikoetxea and M. Uribe-Etxebarria, eds., *Theoretical Issues in the Morphology-Syntax Interface*, pages 97–143. Supplement to the International Journal of Basque Linguistics and Philology.
- Harris, Alice C. 1981. *Georgian Syntax: A Study in Relational Grammar*. Cambridge: Cambridge University Press.
- Hewitt, B. G. 1995. *Georgian: A Structural Reference Grammar*. Amsterdam: John Benjamins.
- Klaiman, M. H. 1991. *Grammatical Voice*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. New York: Mouton de Gruyter.
- Mach'avariani, Maya. 1987. *Kcevis gramatik'uli k'at'egoriis semant'ik'a [Semantics of the grammatical category of version]*. Tbilisi: Mecniereba.
- Melikishvili, Damana. 2001. *Kartuli zmnis ughlebis sist'ema [Conjugation of the Georgian Verb]*. Tbilisi: Logos presi.
- Shanidze, Akaki. 1973. *Kartuli enis gramat'ik'is sapudzvebi I. Morphologia. Meore gamocema [Foundations of the Grammar of the Georgian Language]*. Tbilisi: Tbilisi University gamomcebloba [publishing house].
- Tschenkeli, Kita. 1958. *Einführung in die georgische Sprache*. Zurich: Amiram Verlag.

Denominal Verbs and Noun Incorporation: Uto-Aztecan Evidence for a Unified Syntactic Account*

JASON D. HAUGEN

1 Introduction

One of the classic controversies in the study of the morphosyntax of the indigenous languages of the Americas, which is relevant much more broadly and still largely unresolved, is the correct description and analysis of what has been termed “noun incorporation” (NI). In his foundational statement establishing the modern conception of the issue, Sapir remarked that

[t]he term ‘incorporation’ has been much used in discussion devoted to the structure of American languages. Despite the steadily growing mass of American linguistic material, a good share of the data presented in the last

* This research has been partially supported by NSF Project #0446333. All unattributed data from Yaqui included in this paper have been collected during the course of this project. Many thanks go to Maria Amarillas for sharing her knowledge of Yaqui with our research team: Heidi Harley (PI), Mercedes Tubino Blanco, and myself. For useful discussion of the ideas presented in earlier versions of this paper I am grateful to Shannon Bischoff, Andrew Carnie, Heidi Harley, Jane Hill, Eloise Jelinek, Terry Langendoen, Dave Medeiros, Meghan O’Donnell, Mercedes Tubino Blanco, the participants of the Syntax Salon at the University of Arizona, and the audience at TLS 9. Any inadequacies that might remain are mine alone.

few decades being distinctly superior from the point of view of critical analysis to much that served as illustrative material in earlier days, it can not be asserted that the term is always clearly understood or satisfactorily defined. (1911: 250)

This quotation is as apt as ever, even after nearly a century has elapsed since its initial appearance in print, although many subsequent scholars have adopted Sapir's definition of NI. Under the Sapirean conception, NI is posited to be a morphological process, or, in modern generative parlance, a pre-syntactic (or "lexical") process of compounding a nominal stem to a verbal stem (e.g. Mithun 1984, 1986; Rosen 1989). However, in contrast, there are scholars that have maintained that NI is a syntactic process, as evidenced by such phenomena as the "stranding" of modifiers. Sadock (1980) and Baker (1988) account for modifier "stranding" by positing some syntactic level at which modificational material forms a constituent with an incorporated noun, sharing features such as number and/or gender.

A further issue is the relationship of NI to the similar phenomena that we observe when we consider denominal verb constructions (DVCs). Whereas Sadock (1980, 1986) has regarded the obligatorily incorporating verbal suffixes of West Greenlandic as NI constructions, others (e.g. Mithun 1986) have rejected this position on the grounds that such constructions differ in kind from NI (see also Gerdts 1998).

In the present paper I am going to present evidence from the Uto-Aztec languages to relate DVCs to NI in two distinct ways. The first is empirical, in that DVCs look and behave much like NI constructions, as pointed out by Sadock. The second is theoretical, in that Baker's (1988) head-movement account of syntactic NI has been applied to DVCs, even in languages where it is not at all obvious that nominal roots are incorporating into a verbal predicate (e.g. English, in the work of Hale and Keyser 1993, 2002).

This paper is structured as follows. Section 2 provides a brief historical overview of these issues in order to contextualize the subsequent discussion. Section 3 empirically links DVCs to NI, primarily by means of evidence from Hopi, but also from several other Uto-Aztec languages. Section 4 reviews the most articulated argument to date for separating DVCs from NI (Mithun 1986), but I will argue that there are considerations that suggest that the two are not so distinct after all. Section 5 discusses hyponymous direct objects, which is a problematic issue relevant to syntactic accounts of both NI and DVCs, and section 6 resolves this problem by utilizing the notion of Late Insertion. Section 7 concludes.

2 Historical Background

Sapir defined NI as “the process of compounding a noun stem with a verb . . . no matter what the syntactic function of the noun logically is” (1911: 257). In Sapir’s view, morphology and syntax were obviously distinct realms, and from this perspective he observed that “the sacrifice of syntax to morphology or word-building is indeed a general tendency in more than one American language” (257). Further, contra Kroeber (1909), NI “is primarily either a morphologic or syntactic process; the attempt to put it under two rubrics at the same time necessarily leads to a certain artificiality of treatment” (255). Sapir also emphasized that NI constructions were to be distinguished from DVCs, which involve an obligatorily bound affix attached to a nominal in order to derive a verb, rather than a verbal stem that can appear independently of a nominal host, as is the case with NI.

In contrast, Sadock (1980, 1986) argues that NI in West Greenlandic (Eskimo-Aleut) is necessarily syntactic. The derivational nature of Greenlandic verb structures is shown in (1)-(3). From ostensibly nominal roots such as *qimmeq* ‘dog’ and *pe* ‘thing’, possessive verbs can be formed via a morphological process that is very reminiscent of “incorporation”:¹

- | | |
|--|-------------------------|
| (1) <i>Qimmeq-arpoq</i>
dog-HAVE-INDIC-3sg
‘He has a dog’ | (Sadock 1980: 306 [18]) |
| (2) <i>Pe-qarpoq</i>
thing-HAVE-INDIC-3sg
‘He has something’ | (Sadock 1980: 307 [21]) |
| (3) <i>Qimmimik</i> <i>peqarpoq</i>
dog-INST thing-have-INDIC-3sg
‘He has a dog’ | (Sadock 1980: 307 [24]) |

That the verb is intransitive in these examples is seen by noting that it does not have an object agreement suffix, and that the direct object in (3) appears with instrumental (INST) case marking (Sadock 1980). Since Greenlandic has ergative/absolutive case-marking, the direct object of a morphologically transitive verb normally appears in the absolutive case, as in the following example:

¹ All data in this paper are given with the transcription and glosses of the cited sources. Cases of particularly interesting and/or relevant glosses are discussed in the text.

- (4) *Arnaq takuvara* (Sadock 1980: 304 [2])
 woman(ABS) see-INDIC.-1sg/3sg
 'I saw the woman'

An incorporated nominal can be modified by adjectives, numerals, and the like, which appear in instrumental case just like the direct object in (3):

- (5) *Kusanartumik sapanngarsivoq* (Sadock 1980: 307 [27])
 beautiful-NOM-INST bead-get-INDIC-3sg
 'He bought a beautiful bead.'

The grammatical number of the incorporated nominal is indicated by the external modifiers. Sadock (1980) argues that there are two options for accounting for the modification of the incorporated nominal: either the grammar must assign number twice (onto a pre-formed N-V complex and also onto the modifiers), or only once: onto some modifier-noun constituent. Opting for the latter option, Sadock (1985, 1990) develops an autosegmental theory of morphology-syntax interaction in order to account for these (and other) examples of syntactic word formation.

Mithun (1986) rejects the idea that the Greenlandic data are relevant to the consideration of the cross-linguistic properties of NI, on the grounds that such data involve affixes (i.e. elements that are necessarily bound) rather than being NI constructions proper; i.e. the Greenlandic data involve DVCs. Whereas NI is a discourse-functional option available to some languages, where two roots that can occur independently are compounded for discourse purposes, DVCs involve verbal elements that are necessarily bound. According to Mithun's view, these processes are not the same thing and should not be conflated. Thus, Sadock has not shown that NI is a syntactic process and Sapir's view of NI as a morphological (lexical) process of compounding may be maintained.

Mithun does not directly argue against Sadock's claim that DVCs are syntactic, merely that DVCs do not contain noun-incorporating verbs. If DVCs are in fact syntactic, and if we can show that they are actually incorporating verbs (or at least that they have the same distributional properties as incorporating verbs), then it follows that incorporating verbs are (or at least can also be) syntactic. The distinctions between the two construction types would have to be teased out by examining a language that robustly exhibits both phenomena within the same grammar.

One such language is Hopi, a member of the Northern branch of the Uto-Aztecan language family. Our research has revealed no discernable differences in the syntax of these two construction types in that language. Perhaps most interestingly, both of these construction types appear in a

construction that poses an identical puzzle for purely syntactic accounts of word-formation. In both NI and DVCs, a verbal element with an incorporated nominal can take hyponymous direct object nominals. Such constructions are identified by Mithun (1984) as Type IV or “classificatory” NI.

In section 3 we will examine the relevant data from Hopi and other Uto-Aztec languages that suggest that DVCs are not so different from NI constructions after all, and then in section 4 we will examine point by point Mithun’s arguments for distinguishing these two constructions from each other. I will argue that the most perspicuous account of these constructions is the unified syntactic account provided in sections 5 and 6.

3 The Uto-Aztec Data

3.1 DVCs and NI in Hopi

In a recent paper discussing the extensive data collected by the Hopi Dictionary Project (HDP 1998), K. Hill (2003) exhibits a range of facts that suggest that DVCs and NI constructions have the same syntactic properties in this language. In this section I will briefly review some of this evidence.

The crucial criteria for this discussion are the following: modifier-stranding; hyponymous objects; and the introduction of definite referents.

3.1.1. NI in Hopi

As mentioned above, with true NI constructions, both the verb and the noun can appear as free roots. Such constructions illustrate modifier-stranding in Hopi, as in (6) and (7):

(6) *Naat itam pu-t qatsi-yese.* (K. Hill 2003: 234 [98])
 still we that.-ACC life-sit(PL)
 ‘We are still living that life.’

(7) *Nu’ pay i-t töövu-t aw qötö-tpe.* (K. Hill 2003: 234 [99])
 I well this-ACC embers-ACC to.it head-roast
 ‘I roasted this head over the embers.’

These constructions can also appear with hyponymous objects, where the incorporated nominal acts as a classifier of the external nominal:

(8) *Nu’ yöypala-t küy-tàngta.* (K. Hill 2003: 237 [117])
 I rainwater-ACC contained.liquid-put.into.container(s)
 ‘I put the rainwater into some containers.’

They can also introduce definite referents, a typologically unusual property of NI (cf. Mithun 1984):

- (9) *Nu'* ***pakiw-maqto-ni***; (K. Hill 2003: 241 [143])
I ***fish-go.hunting-FUT***;

noqw itam pu-t enang nööñösa-ni.
so we ***that-ACC*** in.addition.to eat(PL)-FUT

'I'm going fishing, so we can eat it (fish) along with the other food.'

3.1.2. DVCs in Hopi

Unlike NI constructions, DVCs involve a verbal element that is necessarily bound (i.e. that cannot appear without being attached to a nominal root). However, in Hopi DVCs share similar properties to NI constructions, including the appearance of "stranded" modifiers, as in (10) and (11); hyponymous objects, as in (12); and the introduction of definite referents as in (13):

- (10) *Hak i-t kis-ta?* (K. Hill 2003: 234 [96])
who ***this-ACC*** shade-CAUS
'Who built this shade?'

- (11) *Um qa hii-ta ho-'y-va?* (K. Hill 2003: 234 [97])
you not ***what-ACC*** arrow-POSS-INGR
'Didn't you bring any arrows?'

- (12) *Pam tsiili-t nakwa-'y-ta-ngwu.* (K. Hill 2003: 236 [112])
he ***chile-ACC*** feather.worn.on.head-POSS-DUR-HAB
'He (the Hehey'a kachina) wears chili pepper on his head.'

- (13) *Pay kur puma hin'e'wakwmu-y kong-mu-'y-yüñg-qe*
well evidently they ugly(PL)-ACC ***husband-NSG-POSS-DUR(PL)-when***

yaw puma oovi kong-hehep-ya. (K. Hill 2003: 241: 144)
QUOT they therefore ***husband-be.looking.for-PL***

'Since they had ugly husbands, they were looking around for (new) husbands.'

These data suggest that since DVCs and NI constructions share syntactic properties, they should also be amenable to a unified account, at least for Hopi. We now consider denominal verb and NI constructions within the larger context of the Uto-Aztecan languages more broadly.

3.2 Denominal verbs in other Uto-Aztecan languages

Haugen (to appear) reviews the denominal verb morphemes across the Uto-Aztecan family, and finds that the constructions in which they appear typically allow for the stranding of modifiers, as in the following examples:

- (14) Northern Paiute (Snapp et al. 1982: 28)
usu hīci puku-ga-si iwa puku-dua
 he few horse-HAVE-SUBR many horse-INCEP
 ‘He had just a few horses; then his horses became many.’
- (15) Tümpisa Shoshone (Dayley 1989b: 40 [89])
Satü so’oppüh paani’amitu’ih
 that much bread-MAKE-will
 ‘She’ll make a lot of bread’
- (16) Gosiute Shoshone (Miller 1996: 706)
Nī wahatti kahnikantm
 I two-OBJ house-HAVE
 ‘I have two houses’
- (17) Comanche (Charney 1993: 205 [78])
 [suR]ise nīkīnu tammati sonipIφai]
suti=-se nī-kīnu tamma=-ti soni=-pīh-pai
 that.one-CNTR my-father’s.father a.lot-OBJ grass, hay-ABS-pai
 ‘My grandfather had hay (for horses)’
- (18) Southern Paiute (Sapir 1930: 134)
wa’q-utcanI qava’x:A
 two-OBJ-preterite-I horse-GET
 ‘I received two horses’
- (19) Tohono O’odham (Saxton 1982: 141)
n-t wo ha’i kii-ki-t
 I-TNS FUT some RED-house-MAKE
 ‘I’m going to build some houses’

- (20)
- Southeastern Tepehuan
- (Willett 1991: 64 [141])

day ma'n tu-sa'ua-h'iñ
 only one OWN-blanket-1s
 'I have only one blanket'

- (21)
- Western Tarahumara
- (Burgess 1984: 28)

bilé mahtá-ga-me
 one metate-STAT-PRTC
 'I have one metate'

These constructions also typically allow for the appearance of hyponymous objects:

- (22)
- Yaqui

a. *vempo 'uka kari-ta teopo-k* (Jelinek 1998: 217 [61a])
 3pl DET:ACC house-ACC church-PERF²
 'They use that house as a church' (*lit.* 'They church-have that house')

b. *'uka 'ili chu'u-ta nee vuk-ek*
 DET.ACC little dog-ACC 1.sg.NOM pet-PERF
 'That little dog is my pet' (*lit.* 'I pet-have that little dog.')

c. *uu uusi uka chu'u-ta kava'e-k.*
 DET child DET:ACC dog:ACC horse-PERF
 'The child has/uses the dog as a horse'

- (23)
- Tümpisa Shoshone
- (Dayley 1989b: 91 [123])

Nümmü so'oppüh putish pungkupaimmippühantü
 we(exc) many burro pet-HAVE-HAB-PAST
 'We used to have many burro pets'

- (24)
- Gosiute Shoshone
- (Miller 1996: 706)

Isapaippih sukka ponaiha taipai
 Coyote that-OBJ Mouse-OBJ brother-HAVE
 'Coyote has Mouse for a younger brother'

² "PERF" in these examples indicates the perfective morpheme. In Yaqui, any tense, aspect or mood marker can attach to a nominal in order to indicate possession. These constructions have been analyzed as containing a null verbal predicate (Jelinek and Escalante 1988, Haugen 2004).

- (25) Cupeño (Jane H. Hill, p.c.)
túku = 'ep *ne-* 'ásh-lyu *awá-l-i*
 yesterday=r 1s-pet-VB dog-NPN-O
 'Yesterday I had a dog'

With respect to NI with potentially free verb stems, the Uto-Aztec languages vary. It does not appear that many of the Uto-Aztec languages allow for productive NI synchronically, although productive NI is attested within the family, e.g. in Hopi (K. Hill 2003) and in Nahuatl (Merlan 1976). These languages also vary with respect to the synchronic productivity of compounding more generally. Langacker (1977) argues that "it is fair to assume provisionally that the most widespread contemporary UA compounding patterns probably reflect at least approximately the range of major patterns found in PUA" (71), and by the same logic NI can be plausibly reconstructed for the protolanguage.

One potential difference between NI and DVCs is the intransitivizing effect of NI in some languages. Intransitive N-V compounds have been referred to as "Type 1" NI by Mithun (1984), and was what Sapir (1911) had in mind in his discussion on this topic. Jelinek (1998) shows that Yaqui NI constructions differ from typical DVCs in the language, in that NI verbs are intransitive, thus they do not allow modifier-stranding (cf. 26c):

- (26)a. *aapo maaso-ta* *peu-ta-k* (Jelinek 1998: 213 [48])
 3sg deer-ACC butcher-TRAN-PERF
 'He butchered a deer'
- b. *aapo maaso-peu-te-n* (Jelinek 1998: 213 [48])
 3sg deer-butcher-INTR-PAST
 'He was deer butchering'
- c. **aapo bwe'uu-k maaso-peu-te-n* (Jelinek 1998: 213 [48])
 3sg big-ACC deer-butcher-INTRAN-PAST
 'He was [big deer]-butchering'/'He was deer-butchering a big one'

Such NI constructions are not fully productive in Yaqui, which has only a few verbs that can incorporate only a few nouns. However, there are examples of similar cases of NI being used transitively, as in the following examples:

- (27) *Irene panim am-hoo-ria*
 Irene bread-PL 3.PL-make-APPL
 ‘Irene is making bread for them.’
- (28) *Irene am-pan-hoo-ria*
 Irene 3.PL-bread-make-APPL
 ‘Irene is making bread for them.’
- (29) **Irene pan am-hoo-ria*
- (30) *Irene oficiom sii kiam pan-hoo-ria*
 Irene oficio very delicious-PL bread-make-APPL
 ‘Irene is making very delicious bread for the ceremonial officials’

There may be some dialect variation in the domain of transitivity of NI in Yaqui. Null-headed modifiers are otherwise attested in Yaqui,³ so data such as those in (26c) indicate that the issue here truly is one of transitivity. This raises the question of whether we are observing not a distinction between NI and DVCs, on the one hand, but perhaps rather two different kinds of NI (i.e. transitive vs. intransitivizing), on the other. Since the examples from Hopi above are clearly transitive we conclude that the latter is the case.

Rosen (1989) suggests exactly this distinction as the relevant typological property of NI constructions. Whereas Mithun’s categories of NI are discourse-functional, Rosen’s are purely syntactic. Rosen proposes the categories of *compound NI*, which alters the argument structure of the verb by deriving an intransitive verb (as we see in the Yaqui example in 26c), versus *classifier NI*, which creates an N-V compound that retains its transitivity (as in the previous examples from Hopi). Classifier NI is the label applied to NI with both “modifier-stranding” (or, in Rosen’s framework, “null-head” modifiers) and NI with hyponymous objects (which Rosen refers to as “doubling”).

Rosen’s theory of NI is lexicalist, in that the process of compounding the nominal to the verb stem occurs in the lexicon, prior to syntax. With compound NI an intransitive verb is inserted, and with classifier NI a transi-

³ Some examples of Yaqui transitive sentences with null-head modifiers include the following:

- (i)a. *aapo 'uka vicha-k* (Jelinek 1998: 212 [46a])
 3sg DET-ACC see-PERF
 ‘He saw that [one].’
- b. *aapo siali-k vicha-k* (Jelinek 1998: 212 [46b])
 3sg green-ACC see-PERF
 ‘He saw a green [one].’

tive verb is inserted. In section 5 I will discuss a non-lexicalist syntactic approach to these phenomena. Before doing so, however, I will first review Mithun's arguments against including DVCs in the class of NI constructions, in order to justify my discussion of DVCs as incorporating constructions in Uto-Aztecan.

4 Connecting DVC's to Noun Incorporation

The central thesis of Mithun (1986) is that while NI constructions are "very nearly syntactic", they are *not quite* syntactic. Some similarities between NI and syntax proper are that NI manipulates morphological objects (constituents) that are normally related syntactically (i.e. nouns and verbs); it is very productive, unlike some other derivational processes (e.g. affixation); and it involves open class morphemes.

However, NI also shares some properties with morphological word-formation rules, such as the fact that it is highly productive but not completely free; it creates "automatic candidates for lexicalization"; and relatedly, the lexical items formed via NI represent conceptual units, i.e. they are (often) non-compositional. Mithun's points about lexicalization and non-compositionality essentially entail the idea that noun-verb compounds are frequently used to form idioms, such as the following examples from the Northern Uto-Aztecan language Comanche:

(31) Comanche N-V compounding w/ idiomatic usage (Mithun 1984: 855)

- | | |
|--------------------------|--------------------------|
| a. <i>wana-roh-peti-</i> | b. <i>waa-hima</i> |
| cloth-by.force-throw | cedar.tree-take |
| 'to gamble' | 'to celebrate Christmas' |

Mithun's position on this issue reiterates Sapir's stance that morphology is a module of grammar that is strictly distinguishable from syntax.

However, much modern work in morphological theory has blurred the distinction between morphology and syntax, and once this is done Mithun's rationale for separating DVCs strictly from NI largely falls away. Here I will adopt a current theoretical model, Distributed Morphology (DM) (Halle and Marantz 1993, Harley and Noyer 1999), that has placed the issue of the formation of idioms at the forefront in order to show that the attachment of special meanings to linguistic forms is neither an argument for word-hood nor for lexicalism.

DM is overtly non-lexicalist, in that it denies that there is a distinct independent generative mechanism ("Lexicon") that forms words to insert into syntax; rather, word-formation itself is taken to exemplify only one

manifestation of independently necessary syntactic principles. From a decompositional perspective, Marantz (1997) makes the strong claim that even what most speakers of a language might agree are simplex words, such as *cat*, are “phrasal idioms”. From Marantz’s perspective, such words involve roots being interpreted within the context of syntactic configurations involving syntactic functional heads (e.g. “D” or “v”). The basic claim of this approach is that since there cannot be a principled distinction between “syntactic” idioms and “lexical” idioms, there cannot be a principled distinction between “syntactic” word formation and “lexical” word-formation either.

Applying this perspective to the DVC/NI distinction, it appears that Mithun has taken what she assumes to be a lexical process (compounding), and has defined what may or may not be lexical (noun incorporation) in such a way as to presuppose that *it is and can only be* lexical.⁴ She then excludes DVCs from NI by referring to this Sapirean conception of NI, remarking that “it is not entirely clear why one would want to refer to [denominal verb formation] as NI, since it is not obvious what such nouns are incorporated into” (1986: 32).

In order to clarify this issue raised by Mithun, I propose that denominal verb affixes are simply the morphophonological reflexes of the verbal syntactic head “v”. As I will discuss below in section 5, this theoretical link has already been made by Hale and Keyser (1993). In the Hale and Keyser view, denominal verb affixes are “phonologically defective” (and sometimes null) instantiations of v, which unlike typical verb roots actually require the incorporation of the head of their complements in order to be pronounced.

Here I would like to lend a supporting argument for this position based on information inferred from historical change. I assume that *affix-hood* is only one stage in a diachronic process of “grammaticalization”, where free morphemes lose semantic and phonological content over time, undergoing reductions leading to clitic-hood and eventually affix-hood (cf. Bybee et al. 1994, and many others). I would like to suggest that denominal verb morphemes originate as full verbs at v, which over time undergo erosion to phonological defectivity, ultimately requiring the incorporation of the head of their complement in order to be pronounced. However, they remain in the same syntactic location: v. Thus, DVCs are in a sense obligatory NI constructions, and therefore DVCs and NI are amenable to analysis as instantiations of the same (synchronic) syntactic process.

From Mithun’s perspective, however, NI and DVCs are not related in any way, therefore any similarities the two might have would be purely

⁴ Baker (1995) argues that lexical NI co-exists with syntactic NI.

coincidental. Under such a view there is no explanation for the identical characteristics of the two construction types in Hopi, and there is also no reason to expect such similar behavior. In contrast, if the historical relationship of DVCs to NI under the account proposed here is correct, these similarities are not only accounted for, but expected.

In regard to function, Mithun rightly notes that NI constructions are typically optional and have “syntactic analogues” (i.e. verbs can appear in syntactic constructions without incorporating their nominal complement), whereas denominalizing morphemes are always realized as bound affixes that require an incorporated nominal. In Uto-Aztecan, these bound verbal roots also often have syntactic analogues with different, non-bound verbal roots, as with the equivalent ways of expressing the notion of “making” and “possession” in Yaqui, which have both a DVC with a verbalizing affix for making (*-te*) (32a) and possession (*-k*) (33a), as well as a lexical verbs for the equivalent (*hooa*) (32b) and (*hippue*) (33b):

- (32)a. *aapo kari-te*
 3sg house-MAKE
 ‘S/he is building a house’
- b. *aapo kari-ta hooa*
 3sg house-ACC MAKE
 ‘S/he is building a house’
- (33)a. *aapo chuu’u-ø-k*
 3.sg dog-HAVE-PERF
 ‘S/he has a dog’
- b. *aapo chuu’u-ta hippue*
 3.sg dog-ACC HAVE
 ‘S/he has a dog’

In some cases with possessive verbs, there is no semantic or pragmatic difference between the two constructions. However, *hippue* can only be used for alienable possession, and there is an implication of inalienable possession with some usages of the denominal verb:

- (34) a. *aapo tetā-ta mam-po hippue* (Jelinek & Escalante 1988: 419 [29])
 3.sg rock-ACC hand-in have
 ‘S/he has a rock in his hand’
- b. *aapo mam-po tetā-ø-k* (Jelinek & Escalante 1988: 419 [29])
 3.sg hand-in rock-HAVE-PERF
 ‘S/he has a rock (stuck or embedded) in his/her hand’

Thus, it remains the case that speakers can alternate the use of denominal verbs for some functional reason, although it is not clear that this is ultimately relevant to the syntactic analysis of the construction being used. For example, the alienable/inalienable contrast is also made in the Northern

Uto-Aztecan language Gosiute Shoshone (Numic), which utilizes denominal verb affixes (*-kantin* and *-pai*) for each of those functions (Miller 1996).

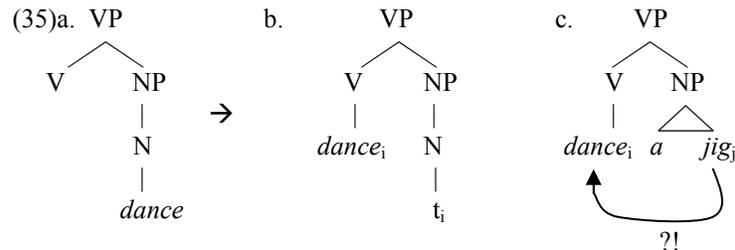
If we are to adopt a theoretical approach such as DM, which minimizes the phonological distinction between free and bound morphemes, might there be some factor that would lead us to maintain that NI and denominal verb constructions are in fact formed by the same process?

I argue in the affirmative. The pertinent factors are identical syntactic distribution, discussed above, as well as an identical semantic selection restriction, to which we now turn.

5 The hyponymous object problem

Whereas Rosen's (1989) lexicalist theory of NI involves noun-verb compounding in the pre-syntactic lexicon, Baker (1988) gives an account wherein NI occurs via head movement ("incorporation") in syntax proper. Like Sadock (1980), Baker supports his syntactic view particularly with evidence from the "stranding" of nominal modifiers, which Baker claims remain in the original position from which the incorporated nominal moves.

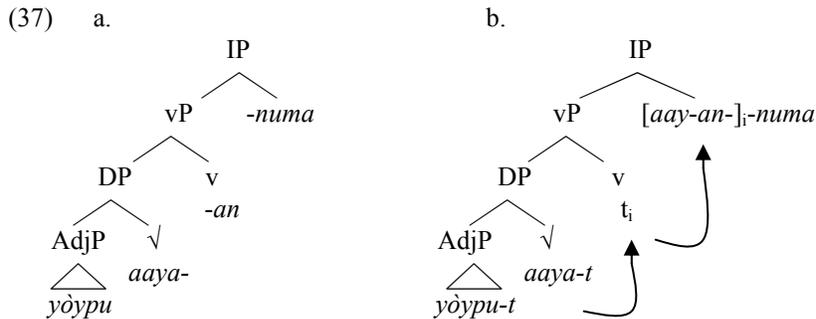
Hale and Keyser (1993) define *denominal* verbs as those verbs formed via incorporation in the technical sense of head-movement, à la Baker (1988). For Hale and Keyser, unergative verbs like English *dance*, *sing*, *saddle*, etc., are formed by the incorporation of a nominal complement into a null verbal predicate ("v"). Their derivation of *dance_v* is shown in (35a-b) below:



(35c) illustrates what I will call *the hyponymous object problem*. Hale and Keyser's syntactic approach does not provide an obvious way for a derived verb to take an overt direct object, if the incorporated nominal originates in the position of the verbal complement. For example, in (35c) the syntactic slot from which the verb *dance* supposedly originates is occupied by non-cognate nominal material, namely, *a jig*.

Cognate objects pose no particular problem for the Hale and Keyser account, assuming that some languages allow for the spell-out of the lower copy of the moved element. Such constructions are attested in Hopi (e.g. 36), and the Hale and Keyser derivation of such is shown in (37).

- (36) *Hak yòy_{pu}-t aaya-t aay-an-numa.* (K. Hill 2003: 239 [131])
 who cracked-ACC rattle-ACC rattle-CAUS-CIRCG
 ‘Someone’s going around shaking a cracked rattle.’



The analysis of such constructions is straight-forward: the spell-out of the lower copy in cognate object constructions is optional in Hopi, but is obligatory in English.

Hyponymous objects as in (35c), on the other hand, initially appear to pose an irreconcilable problem. Hale and Keyser (2002) ultimately abandon their original (1993) approach to denominal verbs, largely because of this problem, and pose instead a separate notion of *conflation*, which is a concomitant of the Minimalist Program notion of Merge, rather than Move.⁵

For the purposes of this paper I will not distinguish the notion of “conflation” from “incorporation”. I will point out that the hyponymous object problem is equally a problem for well-understood cases of syntactic noun incorporation as it is for DVCs under any syntactic account. Since we have seen that DVCs and NI verbs have the same syntactic distribution and observe the same semantic selection restriction (i.e. that any overt nominals in their complement will be interpreted as hyponymous), I will argue that there is ultimately an identical syntactic source for each construction type. It is to the solution of this hyponymous object problem that we now turn.

⁵ Hale and Keyser (1997) attempt to resolve the hyponymous object problem by proposing that movement chains can be broken by “delinking” an incorporated nominal from its base-generated position, via index-deletion. They abandon this approach in their 2002 book, however.

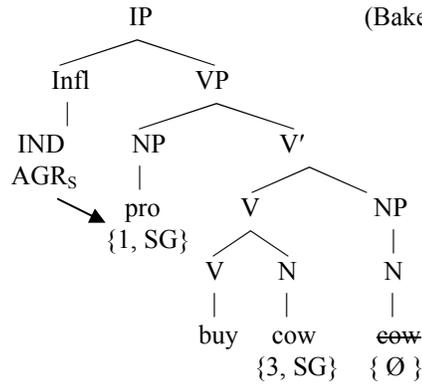
6 The solution: Late Insertion

In the spirit of Baker (1988 *et seq.*), I will present a non-lexicalist approach to NI that involves nominal head-movement in syntax. Within the theoretical framework of DM I offer a novel approach to the hyponymous object problem by utilizing the notion of *Late Insertion*.

In contrast, Baker et al. (2005) propose a *PF-deletion* account of NI. Following the “copy theory” of movement (Chomsky 1995), they claim that languages may vary with respect to whether or not features in copies of a movement chain may be deleted. They assume that NI normally involves the deletion of the lower copy of the moved element (“trace” in earlier theories), which gives the appearance of “stranding” any modifiers in the lower NP:

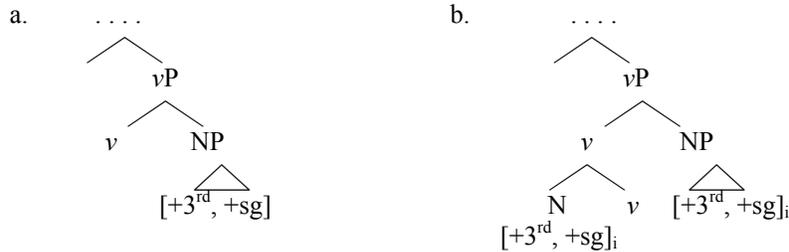
(38) NI with PF-deletion

(Baker et al. 2005: 158 [Fig.2])



With Late Insertion, however, no morphophonological material is inserted into syntactic structures until after all movement operations have already applied. The merge operation merely adjoins a copy of the nominal feature structure of the complement N into the verbal position at *v*:

(39) NI with Late Insertion



NI constructions have three logical possibilities for the spell-out of the complement to the derived N-V complex. They can spell out cognate material, as in (37b) above, or in such English cognate object constructions as *sleep the sleep of the dead*. A movement account of cognate objects in DVCs is supported by denominal verbs in Hindi, where some of these verbs actually require the spell-out of their complement, as in such examples as *khaana khaa-* ‘food-eating’ and *gaanaa gaa-* ‘song-singing’, etc. (Klaiman 1990).

Secondly, these constructions can be spelled out with non-cognate nominal material, which will be interpreted as a hyponymous object, as we have seen with NI constructions in Hopi and denominal verbs across Uto-Aztec. Or, finally, they might allow for non-insertion, which would give the appearance of “stranded modifiers”. The non-spell-out of nominal material presumably depends on the independent availability of null-headed modifiers in the language in question; e.g., English does not typically allow for the appearance of null-head modifiers without a good deal of pragmatic context being available.

As with all syntactic accounts of NI, the analysis proposed here raises the question of why languages may vary with respect to the possible types of complementation that their NI verbs allow. As Baker et al. (2005) point out, Rosen’s lexicalist theory predicts that any hyponymous object should be allowed in languages with Classifier NI, but such is not the case. Southern Tiwa (Kiowa-Tanoan) allows for modifier-stranding, but not cognate or hyponymous objects; Mohawk (Iroquoian) allows for the stranding of the full range of nominal modifiers, but Mapudungun (Araucanian) only allows for the stranding of possessor NPs.

Although modifier-stranding has been taken to be the crucial piece of evidence that NI occurs as a syntactic process (Sadock 1980, Baker 1988), here I adopt Rosen’s claim that the availability of the different types of complementation follows independently from NI itself, as there are in fact languages that allow for null-head modifiers even without NI; this is in fact typical in Uto-Aztec. Thus, I assume that the different possibilities in complementation with NI reflects language-specific factors external to NI. Baker et al. (2005) explain the Mapudungun facts by arguing that possessors (but not other modifiers) appear in argument rather than adjunct positions in that language, which I take to support this position.

Under the lexicalist view, though, the selectional restriction that the incorporated nominal holds over the direct object nominal (i.e. that the latter must be either cognate or hyponymous) must be stipulated. In addition, there is no natural way to rule out the possibility of agent incorporation, which is accounted for by means of the higher structural position of agents

under the syntactic account. Therefore it seems to me that it is preferable to maintain the syntactic account, if it is tenable.

Under the account that I have proposed here, the sharing of features between the incorporated element and its complement occurs because those feature bundles are copies of one another, which, in some cases, can be spelled out with different material (i.e. a classifying nominal root and a hyponymous object nominal root).

There are two major issues for a Late Insertion theory of NI. The first involves the question of how different roots can be inserted to spell out the same features (6.1.); the second is why there is a hyponymous interpretation of the complement NP (6.2).

6.1. Different roots to spell out the same features?

With respect to the issue of how different roots can be inserted into a syntactic tree to spell out the same features, we must consider the nature of the terminal nodes created by syntactic structures. Harley and Noyer (2000) propose two different kinds of syntactic terminal nodes for the post-syntactic insertion of Vocabulary Items. The first are *f-nodes* (\approx functional nodes), for which the spell-out is determined entirely by the syntactic features present (e.g. tense). The second are *l-nodes*, for which a variety of roots may be licensed in any given syntactic environment. It is only in the former that Vocabulary Items compete for insertion, and where the Elsewhere Principle demands that the most highly specified possible candidate be inserted. For l-nodes, however, no such competition takes place.

Under the assumption that the spell-out of l-nodes is non-deterministic and that the encyclopedic information attached to roots is irrelevant to syntax, it follows directly that different roots (e.g. \sqrt{tsiili} and \sqrt{nakwa} , cf. 12) can be inserted to realize the features for which they are licensed (e.g. [+3rd, +sg]).

6.2. Why a hyponymous interpretation?

The second issue that arises under the account proposed here is the (possibly universal) hyponymy relation that is imposed by an incorporated noun to its direct object nominal. I propose that this is a pragmatic matter, akin to Grice's Maxim of Quantity. Since the verb already contains the patient relation in the incorporated noun, any non-cognate root that is spelled out in the lower copy is understood to be giving further specification about that incorporated noun. A hyponymy reading can be coerced even in situations of

novel usage, such as the Hopi example above (12) which involves the classification of a chili pepper (\sqrt{tsiili}) as a member of the set of things that are feathers to be worn on the head (\sqrt{nakwa}), or the Yaqui example (22c) which involves the classification of dogs ($\sqrt{chuu'u}$) as members of the set of animals that can be kept and used as horses ($\sqrt{kava'i}$). The prediction here is that semantic reversals will occur if the relevant nominal roots are reversed. For example, one could plausibly keep horses as dogs or use feathers-to-be-worn-on-the-head as a spice in lieu of chili peppers, although the pragmatics of these situations make such statements odd.

7 Conclusion

To conclude, this paper has related DVCs to NI by presenting a head-movement analysis that uses Late Insertion to account for hyponymous objects in both construction types. I concur with Rosen (1989) that there are two types of NI, although my proposal is non-lexicalist. Compound NI is simply N-V compounding, as proposed by Sapir, Mithun, Rosen and others. Classificatory NI, on the other hand, is formed via head-movement in syntax, as was proposed by Baker (1988), as are DVCs, as proposed by Hale and Keyser (1993).

The upshot of the analysis of DVCs presented here is that data like those from Greenlandic should not be so easily dismissed from our consideration of the cross-linguistic properties of NI. Among other questions that are raised by this account are such issues as why the DVCs of Greenlandic are morphologically intransitive (unlike what we observed above in Uto-Aztecan), and what the articulation of the historical relationship of DVCs to NI might ultimately entail for our synchronic analyses, and vice versa.

References

- Baker, M. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- . 1995. Lexical and Nonlexical Noun Incorporation. *Lexical Knowledge in the Organization of Language*, ed. U. Egli, P. Pause, C. Scwharze, A. von Stechow, and G. Wienold, 4-33. Amsterdam: John Benjamins Publishing Company.
- Baker, M., R. Aranovich, and L.A. Golluscio. 2005. Two Types of Syntactic Noun Incorporation: Noun Incorporation in Mapudungun and its Typological Implications. *Language* 80: 138-176.
- Burgess, D. 1984. Tarahumara. *Studies in Uto-Aztecan Grammar, vol. 4: Southern Uto-Aztecan Grammatical Sketches*, ed. R. Langacker, 1-149. Dallas: SIL.

- Bybee, J., R. Perkins, and W. Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Dayley, J. 1989. *Tümpisa (Panamint) Shoshone Grammar*. Berkeley: University of California Publications in Linguistics volume 115.
- Charney, J. 1993. *A Grammar of Comanche*. Lincoln: University of Nebraska Press.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Gerdts, D. 1998. Incorporation. *The Handbook of Morphology*, ed. A. Spencer and A. Zwicky, 84-100. Malden, MA: Blackwell Publishers.
- Halle, M., and A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. K.L. Hale and S.J. Keyser, 111-176. Cambridge, MA: MIT Press.
- Hale, K.L., and S.J. Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, K.L. Hale and S.J. Keyser, 53-109. Cambridge, MA: MIT Press.
- . 1997. On the Complex Nature of Simple Predicators. *Complex Predicates*, ed. A. Alsina, J. Bresnan, and P. Sells, 29-65. Stanford: CSLI Publications.
- . 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Harley, H., and R. Noyer. 1999. Distributed Morphology. *Glott International* 4: 3-9.
- . 2000. Licensing in the Non-lexicalist Lexicon. *The Lexicon/Encyclopedia Interface*, ed. B. Peeters, 349-374. Amsterdam: Elsevier Press.
- Haugen, J. 2004. Denominal Verbs in Yaqui. *Estudios en lenguas amerindias: Homenaje a Ken L. Hale*, ed. Z. Estrada, A. Fernández, and A. Álvarez, 229-267. Hermosillo, Son: Editorial Unison.
- . To appear. Denominal Verbs in Uto-Aztecan. IJAL.
- Hill, K.C. 2003. Hopi Denominal and Noun-incorporating Verbs. *Studies in Uto-Aztecan*, ed. L. Barragan and J. Haugen, 215-244. Cambridge, MA: MIT Working Papers on Endangered and Less Familiar Languages #5.
- Hopi Dictionary Project. 1998. *Hopi Dictionary/Hopiikwa Lavàytutuveni: A Hopi-English Dictionary of the Third Mesa Dialect*. Tucson, AZ: University of Arizona Press.
- Jelinek, E. 1998. Voice and Transitivity as Functional Projections in Yaqui. *The Projection of Arguments: Lexical and Compositional Factors*, ed. M. Butt and W. Geuder, 195-224. Stanford, CA: CSLI Publications.
- Jelinek, E., and F. Escalante. 1988. 'Verbless' Possessive Sentences in Yaqui. *In Honor of Mary Haas*, ed. W. Shipley. 411-429. Berlin: Mouton de Gruyter.
- Klaiman, M. H. 1990. The Prehistory of Noun Incorporation in Hindi. *Lingua* 81: 327-350.

- Kroeber, A.L. 1909. Noun incorporation in American languages. *Verh. XVI. Intern. Amerikanisten-Kongr.*, Wien: 569-576.
- Marantz, A. 1997. CAT as a phrasal idiom. Ms, MIT.
- Merlan, F. 1976. Noun incorporation and discourse reference in Modern Nahuatl. *International Journal of American Linguistics* 42: 177-191.
- Miller, W. 1996. Sketch of Shoshone, a Uto-Aztecan language. *Handbook of North American Indians, vol. 17: Languages*, ed. I. Goddard, 693-720. Washington D.C.: Smithsonian Institution Press.
- Mithun, M. 1984. The Evolution of Noun Incorporation. *Language* 60: 847-95.
- . 1986. On the Nature of Noun Incorporation. *Language* 62: 32-38.
- Rosen, S.T. 1989. Two Types of Noun Incorporation: A Lexical Analysis. *Language* 65: 294-317.
- Sadock, J. 1980. Noun Incorporation in Greenlandic: A Case of Syntactic Word-Formation. *Language* 57: 300-319.
- . 1985. Autolexical Syntax: A theory of noun incorporation and similar phenomena. *Natural Language and Linguistic Theory* 3: 379-440.
- . 1986. Some Notes on Noun Incorporation. *Language* 62: 19-31.
- . 1990. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. Chicago: University of Chicago Press.
- Sapir, E. 1911. The Problem of Noun Incorporation in American Languages. *American Anthropologist* 13: 250-82.
- . 1930. Southern Paiute: A Shoshonean language. *Proceedings of the American Academy of Arts and Sciences*: 1-296.
- Saxton, D. 1982. Papago. *Studies in Uto-Aztecan Grammar, vol. 3: Uto-Aztecan Grammatical Sketches*, ed. R. Langacker, 93-266. Dallas, TX: SIL.
- Snapp, A., J. Anderson, and J. Anderson. 1982. Northern Paiute. *Studies in Uto-Aztecan Grammar, vol. 3: Uto-Aztecan Grammatical Sketches*, ed. R. Langacker, 1-92. Dallas, TX: SIL.
- Willett, T. 1991. *A Reference Grammar of Southeastern Tepehuan*. Dallas, TX: SIL.

The Syntax of Agreement in Bantu Relatives

BRENT HENDERSON

1 The Typology of Bantu Relatives

Many linguists have found Bantu relative clauses interesting for two reasons. One is a kind of V2 effect found in relative clauses, but not in main clauses, which in many languages requires inversion of the verb and the subject when a nonsubject argument is relativized (see Demuth and Harford 1999 for an overview and useful discussion). Some aspects of inversion will be dealt with in Section 3 below. The second point of interest, the main topic of this paper, is variation in agreement. Restricting attention to non-subject relatives that involve a complementizer, one finds a three-way typology when examining whether the complementizer and/or verb display agreement:¹

- (1) Type 1: Agreement with the subject and relativized NP
Type 2: Agreement with subject only
Type 3: Agreement with relativized NP only.

¹ Assumptions about what constitutes a relative complementizer (as opposed to a relative pronoun) is not uncontroversial. See Zeller (2004) and Demuth and Harford (1999) for opposing viewpoints. In this paper, I consider what I believe are the least controversial cases possible: relative markers that cannot stand alone as pronouns and are phonologically affixal in nature.

In this section, I offer examples for each of these three types, demonstrating with object relative clauses. Type 1 relatives are exemplified by Shona relatives as well as the Zulu so-called “strategy two” relatives. In (2-3) an affixal complementizer displays agreement with the relativized NP while the verb displays agreement with the subject.²

- (2) Mbatya **dza-v-aka-son-era** vakadzi mwenga *Shona*
 10clothes 10REL-3PL-PST-sow-APP women bride
 ‘the clothes which the women sowed for the bride’
 (Demuth and Harford 1999)

- (3) Inja **e-mfana wa-yi-thenga** *Zulu*
 9Dog 9REL-boy 3SG-9OM-buy
 ‘The dog which the boy bought’
 (Poulos 1982)

In Type 2 relatives, while a complementizer is present, it does not display agreement with the relativized NP. However, the verb does agree with the subject. Zulu “strategy 1” relatives as well as Swati relatives display these characteristics. In (4), the marker /a/ does not display agreement with the relativized NP. Similarly with /la/ in (5).³

- (4) incwadi isitshudeni a-**isi-yi-funda-yo** *Zulu*
 9letter 7student REL-7AGR-9OM-read-RS
 ‘the letter that the student is reading’
- (5) umfati tintfombi la-**ti-m-elekelela-ko** *Swati*
 1woman 10girl REL-10AGR-9OM-help-RS
 ‘the woman whom the girls help’
 (Zeller 2004)

² In all examples numbers are used to indicate the noun class of nominals as well as coindexation with agreeing morphemes. Other abbreviations: REL = relative complementizer; AGR = subject agreement; OM = object marker; RS = relative suffix; IMP = imperfect aspect; PERF = perfect aspect; PST = past tense marker; NEG = negative marker; APP = applicative morpheme.

³ The relative marker /a/ plus the subject agreement affix in Type 2 relatives is usually referred to as by the term ‘relative concord’ in the literature. Often the /a/+AGR composition of this marker is obscured by phonological processes like vowel harmony and coalescence. For example, in (4) the relative concord is spelled out as /esi/ while in (5) it is /leti/ (see Mischke 1998 for the full details on such phonological processes in Zulu). Throughout this paper I ignore these phonological transformations for transparency.

Finally, in Type 3 relatives there is no distinct segmental relative complementizer. Rather, relativization is marked by a high lexical tone on the verb. Moreover, it is the relativized NP that triggers morphological subject-verb agreement on the verb. The verb shows no agreement with the subject, which must be overt and postverbal. Many Central Sub-Saharan African languages including Dzamba and Lingala display this type of relativization.

- (6) imundondo **mú** - kpa - aki omoto *Dzamba*
 5jug 5AGR.REL-took-IMP person
 ‘the jug which the person took’ (Bokamba 1976)

- (7) mukanda **mú** – tind - aki Poso *Lingala*
 5letter 5AGR-send-PST Poso
 ‘the letter that Poso sent’

In the next section I attempt to explain this variation in a principled manner.

2 Explaining the Variation: Principles and Parameters

It is standard to take features that are responsible for relativization and wh-movement generally to reside in C while features associated with inflection reside in T. We might therefore simply take the variation above to reflect the presence or lack of phi-features in C and/or T in the three types of relative strategies. However, the data suggests a more interesting conclusion. Consider the variation observed within Zulu, between the so-called “strategy one” (S1) and “strategy two” (S2) relatives. While S2 relatives (in (8) below) are Type 1, S2 relatives (in (9) below) are Type 2. In at least some dialects, these two strategies are both freely available:⁴

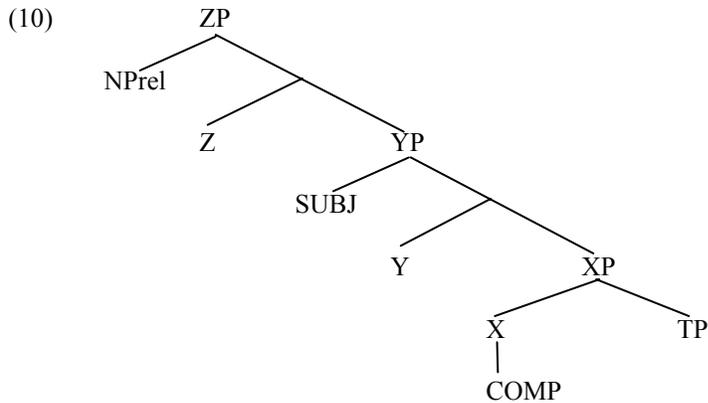
- (8) Inja e-mfana wa-yi-thenga in-hle. *Zulu S2*
 9dog 9REL-boy 3SG-9OM-buy 9AGR-good
 ‘The dog which the boy bought is good.’

- (9) Inja umfana o-wa-yi-thenga-yo in-hle *Zulu S1*
 9dog boy REL-3SG-9OM-buy-RS 9AGR-good
 ‘The dog which the boy bought is good.’

(Poulos 1982)

⁴ Two speakers I have consulted that grew up in metropolitan areas find S2 relatives extremely marginal or ungrammatical. However, two other speakers I consulted who grew up in rural Kwa-Zulu Natal seem to place no restrictions on the use of either strategy.

Not only do these two relatives differ in agreement facts, but also in the position of the relative complementizer /a/. In (8) the complementizer /a/ precedes an overt subject while in (9) it follows the subject, prefixing to the verb. Given (9), if we maintain that relative complementizers like /a/ reside in C, it is necessary to assume a CP domain with more than one projection. This is because both the overt subject and the relativized NP precede the relative complementizer. We therefore require at least three projections in the CP domain: one for the complementizer to head, and two whose specifiers the subject and relative NP can occupy. The required structure appears in (10):

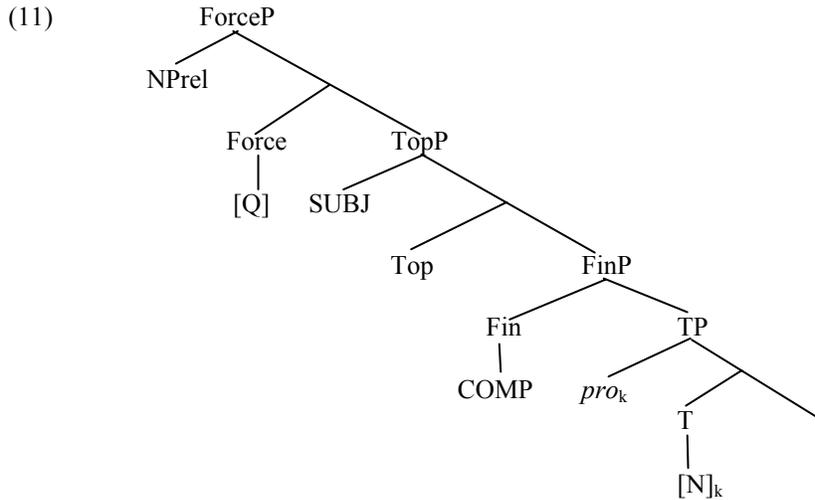


The idea that overt subjects in some Bantu languages are topics residing in the CP domain (rather than structural subjects residing in the TP domain) has been proposed by Letsholo (2002). Under this view, the true structural subject of these languages is not the overt NP itself, but rather a null pronominal *pro* coreferential with the overt subject.

As for the identity of the three projections in (10), those proposed by Rizzi (1997) on the basis of Indo-European languages seem to fit the bill nicely. Rizzi argues that the CP domain should be split up into four possible projections. The topmost projection, ForceP, is responsible for interactions between the clause and the external context, including discourse contexts. It is the locus for clause-typing and clausal operators, including relative operators. The lowest projection, FinP, is responsible for interactions between the CP domain and the inflectional TP domain. Between these two obligatory projections, there are two other projection types which may be present if they are required. One is FocusP (FocP), which serves as a host for focused elements, including wh-question words. FocP is irrelevant for the

present discussion. The other projection is TopicP (TopP), whose specifier hosts left-peripheral topics, including left-dislocated topics.

Rizzi's characterization of the CP domain projections fits nicely with what is required to describe Zulu S1 relatives. The relativized NP resides in SpecForceP while the overt subject resides in SpecTopP. The complementizer /a/ (COMP) I will take to reside in the head of FinP. Placing the set of phi-features associated with subject-verb agreement in T, the structure we arrive at appears in (11)



The structure in (11) characterizes the facts of the Zulu S1 relatives in (9). It also provides a partial solution to the variation we have seen between Zulu S1 and S2 relatives. Recall that unlike in S1 relatives where the relative complementizer appears as a prefix on the verb and therefore after an overt subject, the same marker appears before the overt subject in S2 relatives. This word order difference is easily characterized given the architecture in (11) by assuming that the relative complementizer in Zulu may appear either as the head of FinP or the head of ForceP. In other words, in S1 relatives, /a/ heads FinP while in S2 relatives it heads ForceP.

I would like to note that this possibility of variation in the locus of the relative marker is not antithetical to Rizzi's architectural approach. While it is true that Rizzi associates certain projections with certain interpretation effects, this only affects the interpretation of certain features on the heads of those projections and elements in their specifiers. It does not say anything about the pronounced position of phonological material in the heads of those projections. Consider the now-standard assumption that a parametric

difference (call it the Verb Position Parameter) underlies which syntactic head a verb raises to in natural language. At least since Pollock (1989) it has been argued that in some languages verbs raise to T or AgrS while in others verbs remain in V. Yet there is no interpretation effect associated with this difference. Indeed, the only clear effect it has seems to be pronunciation – a fact that has led some researchers to see head movement as a purely phonological effect (see Boeckx and Stjepanovich 2001, e.g.).

Similar to the Verb Position Parameter, I propose the COMP Position Parameter, the idea that complementizers may reside Force or in Fin.

- (12) COMP Position Parameter: languages differ in whether complementizers reside in Force or Fin.

The parameter in (12) allows for the word order variation in Zulu S1 and S2 relatives. However, it does not account for the variation in agreement facts. Recall that both relatives display subject-verb agreement; however in S1 relatives the relative complementizer /a/ does not agree with the relativized NP while in S2 relatives it does. A rather obvious answer is available: the relative complementizer /a/ carries phi-features when it is realized in Force, but not when it is realized in Fin. But this stipulation alone leaves crucial questions unanswered. Why should the relative strategies differ in this way? Why should the single complementizer /a/ carry agreement in one case, but not in the other?

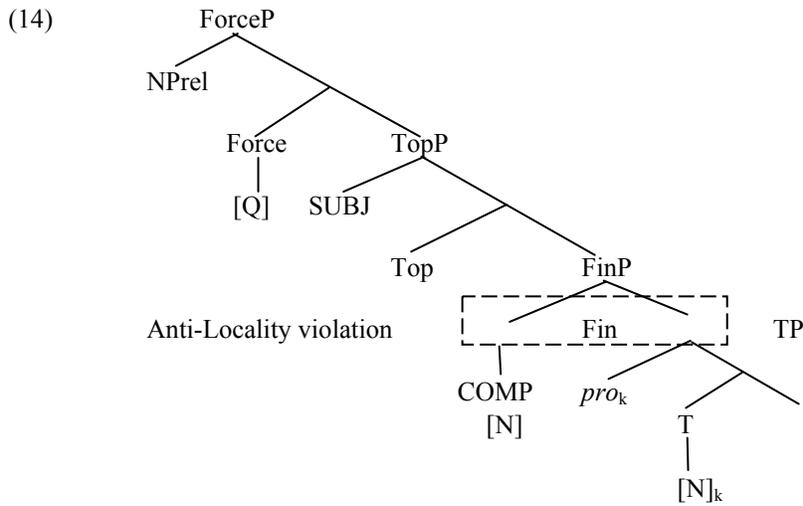
A more explanatory conclusion draws a correlation between the presence of the /a/ marker in Fin and its lack of phi-features. It is this approach that I would like to develop, proposing that a kind of anti-locality effect obtains when two sets of phi-features are “too close” to one another in the syntactic structure. This general concept is explored extensively in Grohmann (2000). Grohmann observes that clause structure has a tri-partite nature, consisting of what he dubs “prolific domains.” Generally speaking, these domains line up with syntactic projections, namely VP (the theta domain), TP (the N-domain) and CP (the T-domain).

Grohmann argues that each domain constitutes a context that is “too local” for movement to occur in. This anti-locality, he argues, prevents movement from, say, SpecTP to SpecAgrS (in the N-domain) or from SpecFinP to SpecForceP (the T-domain). To this general view of things, I would like to add two additional stipulations. The first can be seen as an extension of Grohmann’s concept of anti-locality from the realm of movement phenomena to the presence or realization of phi-features. This is given in (13a). The second is adapted from Rizzi’s (1997) claim that FinP participates both in processes related to the CP domain (hosting A-bar elements,

e.g.) as well as in processes related to the INFL domain (case checking of subjects, e.g.). This is given in (13b).

- (13) a. A prolific domain may contain only one set of N-features.
- b. FinP is a part of both the N-domain and the T-domain

We have thus arrived at a very general set of principles that govern the presence of agreement in the grammar. Combined with the morphological COMP Parameter, these assumptions provide a full characterization of the variation seen between Zulu S1 and S2 relative clauses. Given (13), the fact that the relative complementizer does not display agreement when it occurs in Fin follows. If the complementizer were inserted with its phi-features intact, this would constitute an anti-locality violation, as seen in (14). The grammar's response is to eliminate one set of phi-features, in particular those associated with Fin. Given a derivational system, the latter choice makes sense since T's set of phi-features are introduced first and presumably valued before Fin and its set of phi-features are introduced into the grammar.



From an explanatory perspective, the present account is ideal since it accounts for both the word order and agreement differences between Zulu S1 and S2 relatives with a single morphological parameter, namely that in (12). It must therefore be preferred to an alternative account in which the word order and agreement difference follow from distinct assumptions. Revising

our typology of Bantu relative types to reflect the COMP Position Parameter, we arrive at (15):

- (15) Type 1: COMP in Force
Type 2: COMP in Fin

Note that this morphological variation is limited in the variation it can explain. In particular, it says nothing about the third type of relative we seen, namely Type 3 relatives which display no agreement with a subject, but only with a relativized NP. I repeat an example from Dzamba below:

- (16) imundondo **mú** - kpa - aki omoto *Dzamba*
5jug 5AGR.REL-took-IMP 1person
'the jug which the person took' (Bokamba 1976)

Type 3 relatives cannot be explained by appeal to the COMP Position Parameter. However a similar parameter is possible. Considering that Fin is ambiguously a member of both the N-domain and the T-domain, we can imagine that just as clauses differ with regard to the locus of the features associated with a complementizer (in Force or in Fin), they may also differ with regard to the locus of features associated with inflection. I propose the follow morphological parameter:

- (17) INFL Position Parameter: INFL features may reside in T or Fin.

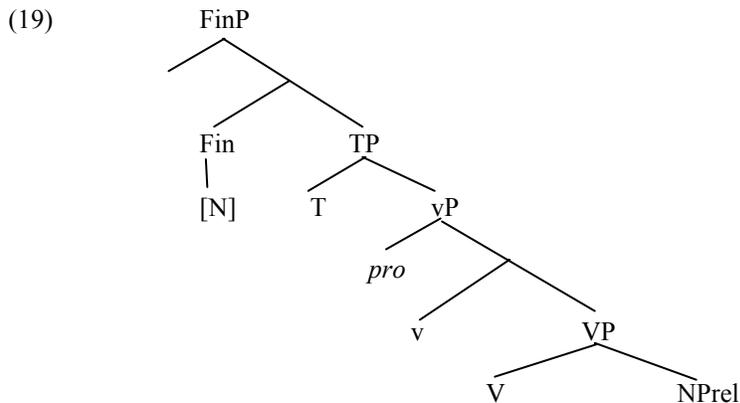
I propose that languages with Type 3 relatives have their INFL features in Fin rather than in T. This morphological difference has two syntactic effects. First, given that Fin is a part of both the N- and T-domains, it follows from the anti-locality assumption in (13a) that no other set of phi-features can be present in either of those two domains. We therefore expect these relatives to agree with one element in the clause, as is the case. This does not immediately answer the question, however, as to why that element must be the relativized NP and cannot be the subject.

The second syntactic effect of having INFL features in Fin is more complex. In Henderson (2006) I argue extensively that languages allowing Type 3 relatives have a single specifier position in the CP domain to which subjects, relativized NPs, and topics must move. This assumption derives the fact that all languages which allow Type 3 relatives also allow topicalization constructions in which a topicalized NPs triggers morphological subject-verb agreement and the subject must be postverbal:

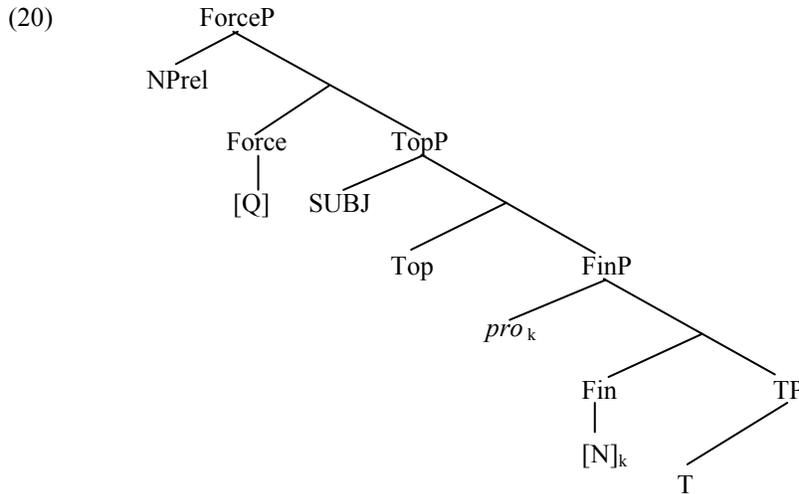
- (18) Imukanda **mu**-tom-aki omwana. Dzamba
 5letter 5AGR-send-PERF 1child
 “The letter, the child sent it.” (Bokamba 1976)

Below, I argue that the option of having INFL features in Fin derives the simplex nature of the CP domain in these languages.

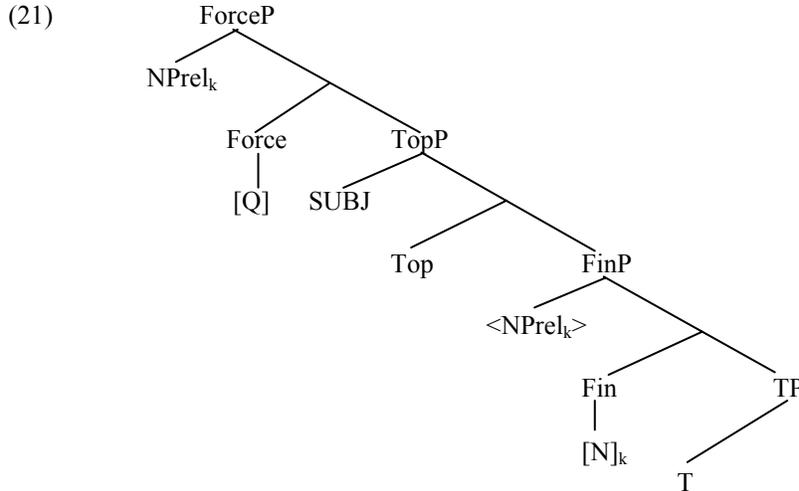
At first glance, the argument languages like Dzamba have a simplex CP layer with only one specifier position would seem to be inconsistent with the view taken here that a complex CP domain like that argued for in Rizzi (1997) is active in Bantu. However, consider the derivation of a Type 3 relative clause. At the point of the derivation in which Fin and its phi-features are introduced, either the subject (here assumed to be *pro*), or the object (represented as NPrel below) may enter a relation with the phi-features and undergo movement to SpecFinP. This situation is represented in (19):



First, let's consider the possibility that the subject moves to SpecFinP. After this movement takes place, the overt subject will be merged in the specifier of a topic position and the relativized NP will undergo movement to SpecForceP. The resulting derivation is represented in (20):



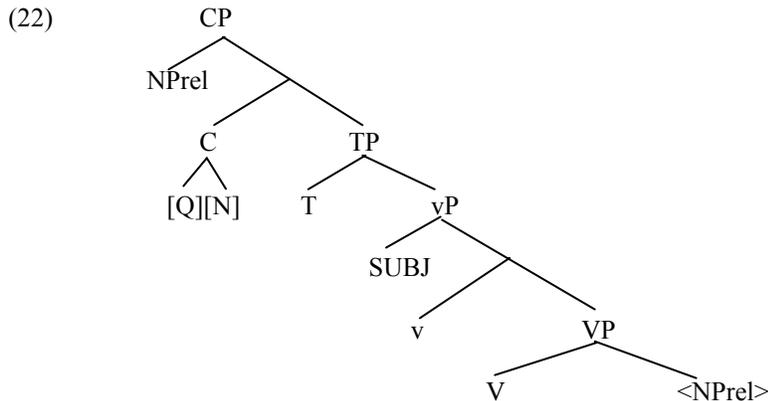
There is a serious problem with the derivation in (20), however. It is standardly assumed that movement must be local. Specifically, when an element moves from one position to another, it may not cross potential landing sites on its way. In (20), the relativized NP undergoing A-bar movement crosses at least one such position, SpecFinP, on its way to SpecForceP. The derivation in (20) is thus ruled out by the general principle of minimality, or Shortest Move. The A-bar movement of NPrel in (20) is not local enough. Now let us consider the other option: namely that the phi-features of Fin in (19) enter a relationship with NPrel and this element moves to SpecFinP. After this movement takes place, the overt subject will be merged in SpecTop and then NPrel will move to SpecForceP. The resulting derivation is represented in (21) where the trace of NPrel is represented as a copy in <>.



The derivation in (21) derives the fact that in Type 3 relatives agreement is with the relativized NP; however, this derivation too suffers from a serious difficulty. Note that in (21) *NPreI* undergoes movement from *SpecFinP* to *SpecForceP*. Ignoring the fact that *SpecTopP* may indeed count as a potential landing spot for purposes of Shortest Move, movement from *SpecFinP* to *SpecForceP* is a clear violation of Grohmann's anti-locality condition on movement since it occurs within the same prolific domain, namely the T-domain. The derivation in (21) is ruled out by this general principle of anti-locality. The A-bar movement of *NPreI* in (21) is too local.

A solution to the problem at hand presents itself once we consider Rizzi's (1997) suggestion that when no *TopP* or *FocP* projections are required between *ForceP* and *FinP*, the latter two projections may collapse into a simplex CP projection. In order for that to occur, of course, the overt subject cannot be merged as a left-peripheral topic, but must be merged in its argument position in *SpecvP*. In that case, *ForceP* and *FinP* can collapse into CP, a single projection whose head will host both the [Q] feature associated with relativization and the [N] features associated with subject-verb agreement. This situation is represented in (22). In (22) *NPreI* has undergone A-bar movement to *SpecCP* where it triggers agreement on the verb. (22) derives the agreement facts of Type 3 relatives as well as the fact that Type 3 relatives require inversion of the verb and subject. In (22) the sub-

ject must remain in-situ since SpecCP, the only site available for movement, is occupied by NPrel.⁵



The derivation in (22) is also easily adapted to represent SVO clauses as well as topicalization OVS clauses in the languages that allow Type 3 relatives. In both cases only one element (the subject in the former case, an object in the latter) will undergo movement from its base position to SpecCP, leaving other elements in situ.

To summarize, assuming that FinP is always an A-bar position, general principles on the locality and anti-locality of movement forces a language to have a simplex CP layer if its INFL features happen to reside in Fin rather than in T. This derives both the agreement and inversion facts of Type 3 relatives. Revising our typology of Bantu relatives once again, we arrive at (23):

- (23) Type 1: COMP in Force; INFL features in T
 Type 2: COMP in Fin; INFL features in T
 Type 3: COMP and INFL features in C.

To summarize, the full range of agreement variation in Bantu object relatives falls out from two simple morphological parameters and general conditions on (anti-)locality. In Type 1 relatives, the two sets of phi-features involved – those associated with COMP and those associated with INFL –

⁵ Another locality issue arises if we assume that phi-features must enter a checking relationship with the nearest c-commanded element with an interpretable set of phi-features, here the subject. See Henderson (2006) for a solution to this problem.

are maximally apart from one another in the clausal architecture, residing in distinct prolific domains. In Type 2 relatives, on the other hand, the COMP features reside in the same prolific domain (the N-domain) as the INFL features since the former reside in Fin. This situation requires COMP to lack a set of phi-features given the anti-locality condition on such features. In Type 3 relatives, the same situation occurs, but in the T-domain. Since COMP and INFL reside in the same domain, only one may have phi-features. Furthermore, since the T-domain has an A-bar character, general conditions on the (anti-)locality of movement rule out a complex CP domain in this situation, resulting in a single CP layer and an obligatorily present in-situ subject.

While deriving the variation in agreement facts in Bantu object relatives, the present work also illustrates an important general characteristic of a particular view of the grammar. It may be that basic morphological variation has no principled explanation. Though some have attempted to explain why in some languages verbs raise to T while in others they remain in V, it is difficult to say how successful these attempts have been.⁶ Similarly, there seems to be no reason other than historical accident why one language would have a complementizer in Force while another would have it in Fin. Yet as I hope to have illustrated here, this kind of essentially random variation may interact with universal syntactic principles such as (anti-)locality and agreement, giving rise to syntactic reflexes that magnify and extend the effect of the minimal underlying variation. Parameters, in this view, are the elements of the syntactic mechanics that cannot really be explained except possibly through the eyes of logical necessity (a verb or a complementizer must be spelled out *somewhere*), yet defining their limits is essential given that the system may react strongly to small morphological variations.

3 Two Kinds of Inversion

In this section, I contrast the inversion seen in Type 3 relatives with the inversion sometimes seen in Type 1 relatives, concluding that the two have different sources in the grammar. Recall that in Type 3 relatives the subject is in-situ since it cannot raise to SpecCP. Therefore, it must be overt and does not trigger agreement on the verb:

⁶ I am thinking in particular of the so-called Rich Agreement Hypothesis which states that the richness of morphological agreement on a verb derives the (im)possibility of verb movement. See Bobaljik (2001) for discussion and arguments that rich agreement reflects syntactic structure and not vice versa.

- (24) Ibitabo **bi** – a – some abana *Kirundi*
 8books 8AGR-PST-read:PERF 2children
 ‘the books that the children read’

However, we have seen that some Type 1 relative clauses, such as those in Shona, also display verb-subject order in object relatives.

- (25) Mbatya **dza-v**-aka-son-era vakadzi mwenga *Shona*
 10clothes 10REL-3PL-PST-sow-APP women bride
 ‘the clothes which the women sowed for the bride’

While inversion in (24) is forced by the syntactic derivation of Type 3 relatives, this is not so in Type 1 relatives like in Shona. In (25) the relative complementizer and its phi-features must be in Force while INFL phi-features are in T as argued above. This seemed obvious in Type 1 relatives in Zulu where the relative complementizer precedes and cliticizes to an overt subject:

- (26) Inja e-mfana wa-yi-thenga in-hle. *Zulu S2*
 9dog 9REL-boy 3SG-9OM-buy 9AGR-good
 ‘The dog which the boy bought is good.’

Why, then, do Type 1 relatives in Shona require inversion while those in Zulu do not? I propose that the answer has to do with morpho-phonological considerations. It is simply a fact about Zulu that relative complementizers can affix either to verbs (if no overt subject is present) or to subjects. The relative complementizer in Shona, on the other hand, does not have this property; it has a morphological requirement that it must affix to the verb. Since the complementizer is in Force in the syntax and the verb is in T, this affixation takes place at the postsyntactic level by PF Merger (Bobaljik 1995). In this view, linguistic elements that are independent syntactically may merge postsyntactically, forming cohesive phonological words. The one requirement for this process to take place is string-adjacency of the elements’ phonological material.

Consider now the derivation of a Type 1 relative in Shona. In (27) the relative complementizer, agreeing with the relativized NP, is in Force while the verb is in T. The overt subject, however, intervenes between the two of them in its location in SpecTopP. There is therefore no way that COMP and the verb can undergo PF Merger and satisfy the morphological requirement that COMP must be a verbal prefix.

- (27) [_{ForceP} Mbatya dza- [_{TopP} vakadzi [_{FinP} [_{TP} *pro* v-aka-son-era...]]]]
 10clothes 10REL women *pro* 3PL-PST-sow-APP
 ‘the clothes which the women sowed’

Fortunately, there is a way out. Both Bobaljik (1995) and Boskovic (2001) argue that, under certain circumstances elements that have raised in the syntax can be pronounced in a lower position from which they have moved. This is clearly a logical possibility allowed by the copy theory of movement – the idea that moved elements leave full copies of themselves in their base position when they undergo movement, rather than a trace. However, this ‘pronounce lower copy’ is clearly a restricted phenomenon since in general elements are pronounced in their moved positions. Interestingly, both authors argue that pronounce lower copy can occur only when the requirements of the morphology/phonology demand it. Returning to (27) we have just such a context. Shona requires that relative markers be verbal prefixes, yet the presence of the subject prevents this from occurring under PF Merger. The solution is that the subject must be pronounced lower in the clause, where it cannot get in the way. I propose that this is the case and that subjects in Shona relatives are pronounced in their base-positions in SpecvP though in syntax (and thus at LF), they reside in the higher position of SpecTP.⁷

If inversion in Type 1 relatives is an instance of pronounce lower copy whereas inversion in Type 3 relatives results from failure to raise the subject from its base position, then the subject will have different LF locations in these two relative types. We therefore expect these two kinds of relatives to display distinct interpretative effects with regard to the subject. Using Type 1 relatives with inversion from Swahili and Type 3 relatives from Kirundi, I present two such effects below, substantiating the claims above.

3.1 Old vs New Information

Though a unified formalization has never been proposed, it is a well-documented fact that postverbal or VP-internal material in Bantu languages receives a new information or focus interpretation (Givón 1972, Bokamba 1976, 1979, Bresnan & Mchombo 1987, Machobane 1987; Demuth & Mmusi 1997). On the other hand, preverbal elements such as subjects tend to be interpreted as old information and function as topics.

⁷ The assumption that overt subjects in SpecTopP are merged there directly and not moved to that position necessitates that in cases of pronounce lower copy the subject be first-merged in SpecvP as the true argument of the clause (rather than *pro*) and then raise to SpecTP.

If the analysis above is on the right track, we should see this difference in the way that subjects are interpreted in pre- and postverbal positions in Kirundi. This prediction is born out. In addition to the Type 3 OVS relatives described above, Kirundi also allows SVO Type 2 relatives. In the latter, the subject may function as old information as seen in (28a). However, in OVS relatives the subject is obligatorily interpreted as new information or as a focused element as indicated in the gloss in (28b). (28b) would be an appropriate response an echo question such as “the books that who read?” or to correct a statement such as “the books that the parents read.”

- (28) a. Ibitabo abana **ba** – a – someye *Kirundi*
 8books 2children 3PL-PST-read:PERF
 ‘the books that the children read’
- b. Ibitabo **bi** – a – someye abana *Kirundi*
 8books 8AGR-PST-read:PERF 2children
 ‘the books that the children read’

In variation between Type 1 relatives with inversion and those without, on the other hand, we do not expect to see this difference between pre- and postverbal subjects since we have claimed that the subject in Type 1 relatives occupies the same LF position in inverted and uninverted structures. This prediction is also born out. Swahili also allows Type 1 relatives with and without inversion. The subject in both the SVO and OVS relatives below has the same interpretation. In fact, the two structures have complete semantic equivalence:

- (29) a. kitabu ambacho mwanfunzi a-li-soma *Swahili*
 7book 7REL 1student 3SG-PST-read
 ‘the book that the student read’
- b. kitabu **a** – li – cho - soma mwanafunzi *Swahili*
 7book 3SG-PST-7REL-read 1student
 ‘the book that the student read’

The interpretation facts thus bear out the analysis from Sections 2 and 3 that though both Kirundi and Swahili display inversion, in the latter it is merely a phonological effect which does not affect semantic interpretation.

3.2 Scope Interpretation

I have claimed that inverted subjects in Swahili raise to SpecTP while in Kirundi they remain within the vP. If this is the case, we expect to see a difference in scope interpretation in case these subjects contain quantifiers. In particular, subjects in Kirundi that remain within the vP should have narrow scope under negation while those in Swahili, which have raised to SpecTP, should have wide scope over negation.⁸ This prediction is born out. The subject in (30a) cannot receive wide scope while the subject in (30b) cannot receive narrow scope.

- (30) a. igitabo nti-gi-a-somye umuntu numwe *Kirundi*
 7book NEG-7AGR-PST-read 1person 1one
 ‘the book that not one person read’
 ‘*the book that one person didn’t read’
- b. kitabu a-si-cho-soma mtu mmoja *Swahili*
 7book 3SG-NEG-7REL-read 1person 1one
 ‘the book that one person didn’t read’
 ‘*the book that not one person read’

These facts substantiate the claim that subjects in Kirundi and Swahili occupy different LF positions.

4 Conclusion

In this paper, I have observed a three-way distinction amongst Bantu object relative clauses with regard to agreement: some relatives display agreement with the subject and the relative NP, some only with the former, and some only with the former. I have argued that this variation arises due to the interaction of general structural and derivation properties of the grammar such as prolific domains and (anti-)locality with a very small amount of morphological variation, namely the locus of COMP and INFL phi-features in the clausal architecture. I have argued that the specifications of the locus of these features for any given language may be essentially random, subject only to historical accident and the logical possibilities afforded by more general principles of the grammar. Finally, I have argued that inversion in Bantu relatives is not a unified phenomenon, (a conclusion also reached in Demuth and Harford (1999)), but may result either from failing to raise a subject as in Type 3 relatives, or as an instance of pronounce-lower copy (as in some Type 1 relatives).

⁸ Thanks to Kyle Johnson for pointing me toward this argument.

5 References

- Bobaljik, J. 1995. Morphosyntax: the syntax of verbal inflection. Doctoral dissertation, MIT.
- Bobaljik, J. 2001. The implications of rich agreement: why morphology doesn't drive syntax. In the *Proceedings of the West Coast Conference on Formal Linguistics* 20:82-95.
- Boeckx, C. and S. Stjepanovic. 2001. Heading Toward PF. *Linguistic Inquiry* 32: 345-355.
- Bokamba, E. 1976. Question Formation in Some Bantu Languages. Doctoral dissertation, Indiana University.
- Bokamba, E. 1979. Inversions as grammatical relation changing rules in Bantu languages. *Studies in the Linguistic Sciences*, 9:1-24.
- Boskovic, Z. 2001. On the nature of the syntax-phonology interface: cliticization and related phenomena. Amsterdam: Elsevier.
- Bresnan, J. and S. Mchombo 1987. Topic, pronoun, and agreement in Chichewa. *Language*, 63: 741-782.
- Demuth, K. & S. Mmusi. 1997. Presentational focus and thematic structure in comparative Bantu. *Journal of African Languages and Linguistics* 18:1-19.
- Demuth, K. and C. Harford. 1999. Verb Raising and Subject Inversion in Bantu Relatives. *Journal of African Languages and Linguistics* 20:41-61.
- Givón, T. 1972. Pronoun attraction and subject postposing in Bantu. In *The Chicago Which Hunt: Papers from the Relative Clause Festival*, ed., P. M. Peranteau, J. N. Levi, and G. C. Phares, 190-97. Chicago Linguistic Society, The University of Chicago.
- Grohmann, K. 2000. Prolific peripheries: a radical view from the left. Doctoral dissertation, University of Maryland College Park.
- Henderson, B. 2006. On OVS in Bantu. Ms., University of Illinois at Urbana-Champaign.
- Letsholo, R. 2003. Syntactic Domains in Ikalanga. Doctoral dissertation, University of Michigan.
- Machobane, M. 1995. The Sesotho locative constructions. *Journal of African Languages and Linguistics*, 16:115-136.
- Mischke, G. 1998. Southern Sotho verbal relative constructions. *South African Journal of African Languages* 18.4: 106-111.
- Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20:365-424.
- Poulos, G. 1982. Issues in Zulu Relativization. Doctoral dissertation, Rhodes University, Grahamstown, South Africa.
- Rizzi, Luigi (1997). The Fine Structure of the Left Periphery. In *Elements of Grammar: Handbook of Generative Syntax*, ed. L. Haegeman, 281-337. Dordrecht: Kluwer.
- Zeller, J. 2004. Relative Clause Formation in the Bantu Languages of South Africa. *Southern African Linguistics and Applied Language Studies*, 22:75-93.

Numeral Classifiers in Lhiimaqalhqama'

SUSAN KUNG

1 Introduction

Like many indigenous languages of Mesoamerica, Lhiimaqalhqama', which is more commonly known as Huehuetla Tepehua, utilizes a system of nominal classification. However, the Lhiimaqalhqama' classifier system is limited to numeral classifiers. A numeral classifier is a morpheme that occurs in the context of a number or quantifier and that classifies or categorizes the referent of a noun based on its inherent properties or its arrangement (Aikhenvald 2000; Allan 1977; Craig 1986; Grinevald 2000). In this paper, I describe the numeral classifier system in Lhiimaqalhqama', I investigate how the system is manipulated by context, and I compare the classifiers with the class of Parts.

Lhiimaqalhqama' is a member of the Totonacan language family and is spoken by fewer than 1,000 people—mostly over the age of forty—in Huehuetla, Hidalgo, in Northeastern Mexico. Lhiimaqalhqama' is a polysynthetic, head-marking (Nichols 1986) language. Sentential word order is pragmatically determined (Mithun 1992), but phrasal word order is more fixed.

The number system of Lhiimaqalhqama' is vigesimal (that is, it is based on the number twenty). The numbers one through thirty-nine are bound morphemes, while the numbers forty and higher are free morphemes, at least in the function of counting. Only the native numbers one to five appear

in my recordings; numbers higher than five are borrowings from Spanish. The majority of native speakers can count only as high as the number ten. I met only one speaker who could count to 111 and another speaker who could count to 100. An informative sampling of the numbers is shown in Table 1.¹

1	tam	21	p'uuxam-tam
2	t'uy	30	p'uuxam-kaw
3	t'útu	31	p'uuxam-kaw-tam
4	t'áti	40	t'u-p'uuxam
5	kiis	50	t'u-p'uuxam-kaw
6	chaaxan	60	t'utum-p'uuxam
7	tujun	70	t'utum-p'uuxam-kaw
8	tzajin	80	t'ati-p'uuxam
9	najatz	90	t'ati-p'uuxam-kaw
10	kaw	100	kiis-p'uuxam
11	kaw-tam	101	kiis-p'uuxam-tam
15	kaw-kiis	110	kiis-p'uusam-kaw
20	p'uuxam	111	kiis-p'uusam-kaw-tam

Table 1: Selected *Lhiimaqalhqama'* Numbers²

¹ All data in this paper come from my field work on *Lhiimaqalhqama'* (Huehuetla Tepehua), which was funded by Fulbright García-Robles, the National Science Foundation (# 0078453), and the Project for the Documentation of the Languages of Mesoamerica. I would like to express my deepest gratitude to the following *Lhiimaqalhqama'* speakers: Nicolás Viguera Patricio and Angela Patricio Tolentino, who helped me with the classifiers, and Antonio Viguera Huerta—may he rest in peace—who taught me the numbers through 111. When don Antonio passed away in October of 2005, so too did the native numbers of the language.

I use a practical orthography; the symbols that differ from the standard IPA symbols are the following: VV = long vowel, C' = glottalized consonant, ' = glottal stop, x = /ʃ/, tz = /ts/, ch = /tʃ/, lh = /l/, y = /j/, j = /h/. The phonemes /y/ and /w/ are semi-vowels. The phoneme /q/ disappeared from the language within the last 60 years, and currently the phoneme /q/ is in the process of merging with the phoneme /ʔ/ (Smythe 2002, 2003). The abbreviations are the following: ART Article, CAUS Causativizer, CL Classifier, EMP Emphasis, FOC Focus, FUT Future, IMPFV Imperfective, IRR Irrealis, LOC Locative, N Noun, NUM Number, PAST Past, PFV Perfective, PL Plural, POS Possessor, PUNC Punctual, Q Question, RPT Reportative, SUB Subject, V Verb, 1 first person, 2 second person, 3 third person, + clitic, ˈ primary accent.

² For the class of numbers, primary stress falls on the last syllable; exceptions bear accent marks. All word-final short vowels are devoiced.

2 Numeral Classifiers in Lhiimaqalhqama'

2.1 Classification of the Classifiers

In Lhiimaqalhqama' the classifiers are prefixed to the numbers and quantifiers, as seen in example (1).

- (1) **maqa**-tam luu
 CL:flexible-one snake
 'one snake'

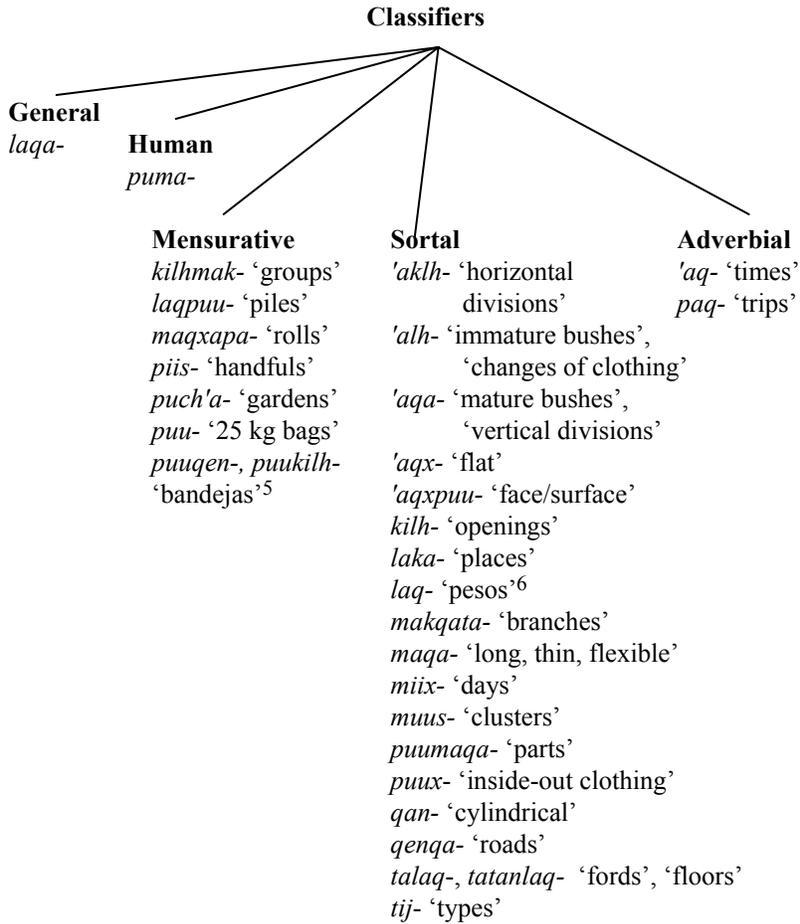
Classifiers are obligatory on the numbers one through thirty-nine and optional on the numbers forty and higher, at least in the task of counting.³ Given that only the numbers one through five appear in my recordings, I cannot know if the classifiers are obligatory on numbers greater than five in contexts other than counting. According to Aikhenvald (2000: 100), it is very common for numeral classifiers to be used obligatorily with small numbers and optionally with larger numbers. The division between 'smaller' numbers and 'larger' numbers is language specific; in the case of Lhiimaqalhqama', the division is between thirty-nine and forty.

Many researchers have made typologies of the classification of nouns, including Aikhenvald 2000, 2004, Allan 1977, and Grinevald 2000. I have not strictly adopted any of their typologies, but rather I have incorporated aspects of each in the categorization of the numeral classifiers in Lhiimaqalhqama'. I have divided the classifiers into five groups: (i) the general classifier *laqa-*, which can be used in place of any of the other classifiers; (ii) the human classifier *puma-*;⁴ (iii) mensurative classifiers that measure entities; (iv) sortal classifiers that identify inherent characteristics of entities; and (v) adverbial classifiers that function only as adverbs. The classification is shown in Figure 1, which is a complete list of the Lhiimaqalhqama' numeral classifiers.

³ The only exception is the use of the number *tam* 'one' without a classifier as the indefinite article.

⁴ There is no specific classifier for animals or inanimate entities.

Figure 1: Classification of the Numeral Classifiers in Lhiimaqalhqama'



⁵ A *bandeja* is a unit of measure in the Mexican market place; it is used to measure and weigh dry, pourable, noncountable entities like coffee beans and black beans.

⁶ The monetary unit of Mexico.

2.2 Morphosyntax of the Classifiers

Classifiers in Lhiimaqalhqama' appear in two morphosyntactic constructions: (i) in adjectival position or as anaphora and (ii) in adverbial position. Although sentential word order is pragmatic, phrase word order is relatively fixed: in particular, (i) nouns follow adjectives and (ii) adverbs occur directly before the verb or at the end of the sentence, and they never appear immediately before a noun.

The first construction in which classifiers occur, is one in which a classified number modifies an optionally overt nominal, as seen below in example (2).

(2) [CL-NUM (N)]

When the noun occurs, the classified number precedes it, as seen in the examples in (3).

- (3) a. **'aqx**-tam 'alhik
 CL:flat-one paper
 'one sheet of paper'
- b. **puma**-t'útu lapanák
 CL:human-three people
 'three people'

When a classifier categorizes its prototypical referent, the noun may be omitted, and the classified number or quantifier behaves anaphorically, as seen in the examples in (4).

- (4) a. taas **puma**-chuux (lapanák) ka-ta-min-aa
 Q **CL:human**-how.many (people) IRR-3PL.SUB-come-FUT
 'How many people will come?' [Q7]
- b. naa **puma**-lhuu ka-ta-min-aa
 EMP **CL:human**-many IRR-3PL.SUB-come-FUT
 'Many people will come.' [Q7]

However, when a classifier categorizes a noun that is not its prototypical referent, omission of the noun changes the meaning of the sentence, as seen in the examples in (5).

- (5) a. **qan-t'uy** lapanák
CL:cylindrical-two people
 'two people'
- b. **qan-t'uy**
CL:cylindrical-two
 'two cylindrical things'
 NOT: 'two people'

Furthermore, when the classifier is *laqa-*, the general classifier, omission of the noun can result in a lack of meaning, as seen in example (6).

- (6) a. **laqa-t'uy** lapanák
CL:general-two people
 'two people'
- b. **laqa-t'uy**
CL:general-two
 'two X'
 NOT: 'two people'

In the second morphosyntactic construction in which a classifier may appear in Lhiimaqalhqama', the classified number occurs in the two adverbial positions, shown in example (7). The first adverbial position occurs immediately before the verb, as seen in (7a), while the second adverbial position occurs clause-finally, as seen in (7b).

- (7) a. CL-NUM V
 b. V . . . CL-NUM

The examples in (8) show a classified number acting as an adverb. In the examples in (8a) and (8b), the adverb precedes the verb, while in (8c), the adverb is clause-final.

- (8) a. **'aq-t'uy+ch** xa-k-masqati-y
CL:times-two+PUNC PAST-1SUB-try-IMPV
 'I tried two times.' [T0069: 005]

- b. **paq**-t'ut'u+ch xa-la-y juu 'atzi'
CL:trips-three+PUNC PAST-can-IMPV ART girl

juu 'ix-xkaan
ART 3POS-water
'The girl went for water three trips (times).' [Q7]
- c. maa-lach'ap'a-y juu 'alhik puu-**aqx**-t'uy
CAUS-glue-IMPV ART paper LOC-**CL:flat**-two
'She glues the paper in two places.' [MNB13: 96]

3 Contextual Dependency

Nouns in Lhiimaqalhqama' are not rigidly divided into different classes, as the nouns in the Bantu languages are; therefore, instead of serving to coreference a noun class, the classifiers are fluid and serve to highlight certain characteristics of the noun that are relevant to a given context.

For example, in (9) the noun is *lapának* 'person'. The classifier in example (9a), *puma-*, is the default human classifier that is prefixed to any number that modifies a human noun. Although the same noun *lapának* also appears in (9b), the classifier *qan-* is used to modify a long, cylindrical noun; here the use of *qan-* instead of *puma-* communicates the fact that the people are lying down, more dead than alive. The example in (9c) also contains the human noun *lapának*, and here it is modified by the classifier *'aklh-*, which indicates that the noun is divided along a horizontal axis. The use of the classifier *'aklh-* indicates that the division between the human half and the animal half is horizontal, not vertical, and this information is more important to the context than the fact that the noun *lapának* is human.

(9) noun = *lapának* 'person'

- a. juu **puma**-tam lapának nii-lh
ART **CL:human**-one person die-PFV
'One person died.' [T0009: 001]
- b. maa ta-laxtaqni-lh+ch juu x-taqanqat-'an
RPT 3PL.SUB-contract-PFV+PUNC ART 3POS-illness-PL.POS

juu **qan**-tam **qan**-tam lapának
ART **CL:cylindrical**-one **CL:cylindrical**-one person
'The people contracted the illness one by one.' [T0057: 019]

c. waa 'aklh-tam lapának
 FOC CL:horizontal-one person

'aklh-tam maqtili'
 CL:horizontal-one wild.animal
 'It is half person, half wild animal.'

[MNB16: 50]

The ten examples in (10) demonstrate the use of various classifiers that describe arrangements or measurements of the noun *k'iw* 'tree'. All of these examples are of sortal classifiers with the exception of examples (10d), which is a Part (see Section 4), and examples (10e) and (10f), which are mensurative classifiers. The examples in (10a) and (10b) use the general classifier *laqa-* and the cylindrical classifier *qan-*, respectively, in order to form the noun phrase 'two trees'. The classifier in (10c) indicates 'types' of trees. In example (10d), *kinka-* is a Part that means 'point' and that is used here as a classifier; in this context, the use of *kinka-* indicates that the two trees are pointy and bare of branches and leaves. The classifiers in examples (10e) and (10f) measure the noun by 'gardens' and 'rolls', respectively. The examples in (10g) through (10j) form a set using the compound noun *lht'a-qala-k'iw*, meaning 'board'. In (10g) the classifier *'aqx-* simply indicates that the board is flat, while the classifiers in (10h) through (10j) indicate the form of the division between the parts of the board. In (10h) *puumaqa-* indicates only that the parts all come from the same board; it does not actually specify information about the shape of the division. The arrows in the drawing in the corresponding Figure 2 indicate that the division can be in any part of the rectangle. This classifier is frequently used with pieces of fruit or bread. The classifier in example (10i) *'aqqa-* indicates that the division in the board is vertical, as seen in the drawing in Figure 3; and the classifier in example (10j) *'aklh-* indicates that the division is horizontal, as seen in the drawing in Figure 4.

(10) noun = *k'iw* 'tree'

a. **laqa-**t'uy k'iw
 CL:general-two tree
 'two trees'

b. **qan-**t'uy k'iw
 CL:cylindrical-two tree
 'two trees'

- c. **tij**-t'uy k'iw
 CL:type-two tree
 'two types of tree'

- d. **kinka**-t'uy k'iw
 CL:point-two tree
 'two pointy trees'

- e. **puch'a**-t'uy k'iw
 CL:garden-two tree
 'two gardens of trees'

- f. **maqxapa**-t'uy k'iw
 CL:roll-two tree
 'two rolls of firewood'

- g. **'aqx**-t'uy lht'aqala-k'iw
 CL:flat-two flat-tree
 'two boards'

- h. **puumaqa**-t'uy lht'aqala-k'iw
 CL:part-two flat-tree
 'two parts of a board' (cut from the same board)

- i. **'aqa**-t'uy lht'aqala -k'iw
 CL:vertical-two flat-tree
 'two vertical sections of board' (cut from the same board)

- j. **'aklh**-t'uy lht'aqala-k'iw
 CL:horizontal-two flat-tree
 'two horizontal sections of board' (cut from the same board)

Figure 2: *puumaqa-*



Figure 3: *'aqa-*

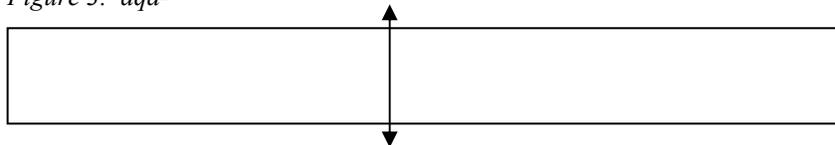


Figure 4: 'aklh-



Finally, the six examples in (11) are based on the noun *kapen* ‘coffee’. The classifiers in examples (11a) through (11e) are sortal, while the one in (11f) is mensurative. Again the general classifier *laqa-* appears with the noun in example (11a). In (11b) *'alh-* refers to an immature plant that is just beginning to grow, while in (11c) *'aqa-* refers to a mature plant that is already fully grown and ready for harvesting. In (11d) *máqgata-* refers to branches of the plant, and in (11e), *muus-* indicates clusters of the fruit of the plant. Given the specificity of the classifiers in (11b) through (11e), it is obvious that plant life (like coffee, bananas, beans, corn, etc.) is very important for the Tepehua culture. The classifier *puuqen-* in (11f) is mensurative and it measures the coffee beans by bandejas.⁷

(11) noun = *kapen* ‘coffee’

- | | | |
|----|--|--------|
| a. | laqa- t'áti | kapen |
| | CL:general-four | coffee |
| | ‘four coffee beans’ | |
| b. | 'alh- t'áti | kapen |
| | CL:immature.bush-four | coffee |
| | ‘four immature coffee bushes’ (just beginning to grow) | |
| c. | 'aqa- t'áti | kapen |
| | CL:mature.bush-four | coffee |
| | ‘four mature coffee bushes’ | |
| d. | máqgata- t'áti | kapen |
| | CL:branch-four | coffee |
| | ‘four branches of a coffee bush’ | |

⁷ See Footnote 5.

- e. **muus-t'áti** kapen
CL:cluster-four coffee
 'four clusters of coffee beans'
- f. **puuqen-t'áti** kapen
CL:bandeja-four coffee
 'four bandejas of coffee beans'

4 Parts and Classifiers

Lhiimaqalhqama' has a class of lexicalized prefixes that refer to Parts (of a body or a structure) and that appear on verbs, nouns, and dimensional adjectives.⁸ The Part *kik-* 'mouth' is prefixed to a verb in (12a), to a noun in (12b), and to a dimensional adjective in (12c).

- (12)a. **kik-ch'aqa-y**
mouth-wash-IMPFV
 'She washes her mouth.'
- b. **kik-ch'awti**
mouth-hair
 'mustache', 'beard'
- c. **kik-lhman**
mouth-long
 'bearded'

While *all* of the numeral classifiers in the related language Papantla Totonac come from the class of Parts (Levy 2004: 284), this is not the case in Lhiimaqalhqama', where only some of the classifiers seem to have origins in the Parts class. In fact, the majority of the Lhiimaqalhqama' classifiers are distinct from the Parts; however, there is overlap—both syntactic, semantic, and phonological—between the Parts and the classifiers.

There are five points of comparison between the Parts and classifiers.⁹ First, some of the classifiers bear no similarities—either syntactic, semantic or phonological—to any of the Parts prefixes. These classifiers are shown in Table 2 and are exemplified in the majority of the examples above.

⁸ Please see the Appendix for a complete list of Parts.

⁹ It is important to note that all of the Parts that contain /k/ or /q/ participate in sound symbolic phonemic alternations that are symbolic of size or affection, e.g. 'aq- 'head' ~ 'ak- 'small head' (Smythe Kung, in press).

Prefix	Part Meaning	Classifier Meaning
'aklh-	---	horizontal division of N
'alh-	---	immature bush; change of clothing
'aq-	---	mature bush
'aqxpuu-	---	face, surface
kihlmak-	---	group of people or animals (e.g. teams, herds)
makqata-	---	branch
maq-	---	long, thin, flexible N
maqxapa-	---	roll of N
miix-	---	day or week
paq-	---	trip
piis-	---	handful of N
puch'a-	---	garden of N
puma-	---	human
puumaqa-	---	part of N
puuqen- ~ puukilh-	---	bandeja or dipperful of N
puux-	---	inside-out clothing
qan-	---	cylindrical N
talaq- ~ tatanlaq-	---	floor of a building; ford of a river
tij-	---	type of N

Table 2: Prefix is a Classifier

Second, some of the Parts prefixes bear no similarities to any of the classifiers. These Parts are shown in Table 3.

Prefix	Part Meaning	Classifier Meaning
'ak-	head	---
ka-	nose, tip	---
lak-	body	---
lakpuu-	face, eye	---
muuntz'a-	forehead	---
tamp'us-	belly	---
tzoqot-	knee	---

Table 3: Prefix is a Part

Third, some of the Parts can be prefixed to numbers in place of classifiers. In these cases, the Part preserves its part meaning, as seen in the examples below in (13).

- (13)a. **ch'an-t'uy**
foot-two
 'two feet'
- b. **katu-t'uy**
ear-two
 'two ears'

However, the constructions in (13) present the following question: Are these prefixes Parts used syntactically as classifiers, or are they classifiers that are semantically and phonologically identical to Parts? For now, I consider them to be Parts, not classifiers, and I do not include them in the list of classifiers in Figure 1. These prefixes are shown in Table 4.

Prefix	Part Meaning	Classifier Meaning
ch'an-	foot, paw	foot, paw
kapii-	palate (of mouth)	palate (of mouth)
katu-	ear	ear
kik-	mouth, beak, edge	mouth, beak, edge
kilhtu-	edge	edge
lakapaa-	head	head
laqxtan-	cheek, jaw	cheek, jaw
laqxtii-	crown of the head	crown of the head
mak- ~ maq-	hand	hand
maqaxtu-	elbow, corner	elbow, corner
muunti-	forehead	forehead
paaka-	armpit, wing	armpit, wing
piixtu-	neck	neck
qaatu-	thigh	thigh
staa-	back	back
tampuu-	belly	belly
tan-	front of trunk of body	front of trunk of body
tasa-	tooth	tooth
tii-	tail, butt, hip	tail, butt, hip

Table 4: Part & Classifier Have Same Meaning

Fourth, other Parts are homophonous with some classifiers, but the Part and the classifier have different meanings, as seen in Table 5. For example,

'aq-, when used as a Part, means 'head', but when used as a classifier, it indicates the number of times the action of the predicate was performed.

Prefix	Part Meaning	Classifier Meaning
'aq-	head	number of times action of V completed
laka-	body	place
laq-	body	peso
laqa-	body	general
laqpuu-	face, eye	pile of N
qenqa-	nose, tip	road

Table 5: Part & Classifier Have Different Meanings

Finally, other Parts are homophonous with some classifiers, and the Part and the classifier have transparently similar meanings, as seen in Table 6. For example, 'aqx- as a Part means 'shoulder', and as a classifier, it means 'flat'. An example is seen in (10d), where *kinka-* is prefixed to the number *t'uy* 'two'.

Prefix	Part Meaning	Classifier Meaning
'aqx-	shoulder, upper back	flat
kilh-	mouth (full N)	opening
kinka-	nose, tip	point, tip (not nose)
muusa-, muus-	groin	cluster
puu-	innards, insides	gunny sack, abt 25 kg

Table 6: Part & Classifier Have Similar Meanings

In conclusion, nineteen of the twenty-nine classifiers shown in Figure 1—that is, two-thirds of the classifiers—do not correspond syntactically, semantically, or phonologically to any of the Parts prefixes. Eleven of the twenty-nine—or one-third—are homophonous with Parts; of those eleven, six are semantically distinct while five are semantically similar. Given that one-third of the classifiers are homophonous with corresponding Parts and that two-thirds¹⁰ of the Parts shown in Table 8 of the Appendix can be used as classifiers, it is probable that at least some of the members of the two classes share a common origin. However, given that the majority of the classifiers are not phonologically or semantically similar to the Parts, it is also possible that these particular classifiers have some other origin.

¹⁰ Nineteen of thirty.

5 Conclusion

This work is a first pass in the description of the classifier system of Lhiimaqalhqama'. Unfortunately, I doubt that I will discover more classifiers because today the majority of Lhiimaqalhqama' speakers use only the general classifier *laqa-* and the human classifier *puma-*. Although I have found some thirty classifiers during the task of elicitation, only eight of these appear in the recordings that I have analyzed so far. These eight classifiers appear in Table 7. I should add that I still have many recordings to analyze, so I hope to find more classifiers, but I am not overly optimistic.

Classifier	Gloss
'aklh-	parts
'aq-	times
'aqx-	flat
kilh-	openings
laq-	pesos
laqa-	general
puma-	human
qan-	cylindrical

Table 7: *Classifiers Found in Recordings*

Although I have begun the task of categorizing the Lhiimaqalhqama' numeral classifiers here, the categorization needs more specification, especially in the area of form; for example, many of the classifiers under the Sortal category, could be further divided based on whether they describe the general shape of a referent (e.g. *maqa-* 'flat, long, and flexible') or whether they describe a more specific type of noun (e.g. '*alh-* 'immature bush').

Additionally, further analysis needs to be done regarding the morphology that can cooccur with the numeral classifiers. For example, in (8c), a locative prefix precedes the numeral classifier; testing still needs to be done to determine what other morphemes may cooccur with the classifiers.

Finally, the relationship between the classifiers and the Parts prefixes needs to be investigated further, along with the question of the origins of the numeral classifiers.

References

- Aikhenvald, A. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Aikhenvald, A. 2004. Nominal Classification: Towards a Comprehensive Typology. *Sprachtypologie und Universalienforschung* 57 (2/3): 105-116.

- Allan, K. 1977. Classifiers. *Language* 53 (2): 285-311.
- Craig, C. (Ed.). 1986. *Noun Classes and Categorisation*. Amsterdam: John Benjamins Press.
- Grinevald, C. 2000. A Morphosyntactic Typology of Classifiers. *Systems of Nominal Classification*, ed. G. Senft, 50-92. Cambridge: Cambridge University Press.
- Levy, P. 2004. Parts in Papantla Totonac and the Genesis of Systems of Numeral Classification. *Sprachtypologie und Universalienforschung* 57 (2/3): 280-299.
- Mithun, M. 1992. Is Basic Word Order Universal? *Pragmatics of Word Order Flexibility*, ed. D. Payne, 15-61. Amsterdam: John Benjamins.
- Mithun, M. 1997. Lexical Affixes and Morphological Typology.' *Essays on language function and language type*, ed. J. Bybee, J. Haiman, and S. Thompson, 357-371. Amsterdam: John Benjamins.
- Nichols, J. 1986. Head-marking and Dependent-marking Grammar.' *Language* 62 (1): 56-119.
- Smythe, S. 2002. The Loss of Uvular Stops in Huehuetla Tepehua. Presentation at the Annual Meeting of the Society for the Study of Indigenous Languages of the Americas, San Francisco, CA.
- Smythe, S. 2003. Reconstructing Lost Phonemes in Huehuetla Tepehua Using 'Affectionate Speech'. *Texas Linguistic Forum* 45: *SALSA 10*, ed. I. Mey, G. Pizer, H-Y. Su and S. Szmania, 167-176. Austin: UT Linguistics Dept.
- Smythe Kung, S. In Press. Simbolismo Sonoro y Lenguaje Expresivo en el Tepehua de Huehuetla. *Memorias del VIII encuentro internacional de lingüística en el Noroeste*, Hermosillo, Sonora, México.

Appendix

Prefix Part	Full Noun	Gloss
'aq-, 'ak-, lakapaa-	'aqtzulh	head
'aqx-, 'akx-, 'aqxp'in-	'aqxp'un	shoulder, upper back, flat surface
'aqxspuu-	'ukxpu'	face, surface
ch'an-, tz'an-	ch'aja'	foot, paw
kapii	kapiiya'	palate
katu-	'aqaxqolh	ear
kik-, kilhtu-, qelh-	kilh	mouth, beak, edge
kinka-, ka-	kinkati'	tip, nose
kinka-	lhiixin	tip, nose
laka-, lak-, laqa-, laq-	lakatunaj	body
lakpuu-, laq(a)puu-, laq-, lak-	'ukxpu'	face
lakpuu-, laq(a)puu-, laq-, lak-	laqchulh	eye
laqx'tii	'akpuux	crown of the head, hair
mak-, maq-	maka'	hand, arm
maqaxtu-	maqaxtu'	elbow, corner
muuntz'a-, muunti	muuntz'an	forehead
muusa-	muusan	groin
paka-	pakapu'	armpit, wing
piixtu-	piixtu'	neck
puu-	---	insides, inside
qaq-, kaq-, laq(a)xtan-	laqxtan	cheek
qaq-, kaq-	kaalhtz'an	jaw
qaatu-, 'aatu-	qaatu'	thigh
staa-	puulakan	back
tamp'us-	tamp'uktz'ulh	belly button
tan-	tampuu	stomach
tan-	tankilhak	chest
tasa-	tatzalat	tooth
tii-	tiimus	tail, butt, hip
tzoqot-	tzoqot	knee

Table 8: *Lhiimaqalhqama' Parts Class*

Argument Binding and Morphology in Chichewa¹

SAM MCHOMBO

1 Introduction

Bantu verbal morphology traditionally comprises a verb root (VR) to which are suffixed extensions such as the causative, applicative, reciprocal, passive, etc. and are prefixed morphemes that encode negation, subject marker and object marker that cross-reference Topic noun phrases, tense/aspect, modality, etc. The latter differ from the suffixes in both form and function. Formally the suffixes have a –VC- structure, as opposed to the canonical CV syllable structure. Functionally the verbal suffixes affect argument structure (cf. Dlayedwa 2002; du Plessis & Visser 1992; Hoffman

¹ Preliminary versions of the material reported here were presented at the workshop on argument structure held at the Center for Advanced Studies in Theoretical Linguistics, University of Tromsø, Norway (Nov. 2004) and at the 3rd annual symposium of The Society of Linguistics Undergraduate Students at The University of California, Berkeley (Nov. 2004). I am grateful to the participants, most especially to John Bowers, Setumile Morapedi, Peter Muriungi, Peter Svenonius, Tara Taraldsen, and Edwin Williams, for stimulating discussion. The ideas have equally been influenced by previous collaboration with Alex Alsina, Joan Bresnan, Mary Dalrymple, Makoto Kanazawa, Pascal Kishindo, Al Mtenje, Armindo Ngunga, and Stanley Peters. None of them necessarily agrees with or endorses the views expressed, responsibility for which rests with the author.

1991; Letsholo 2002; Mabugu 2001; Morapedi 2006). Sentence 1 below illustrates the basic morphological organization of the verb in Chichewa²:

- (1) Mkángo u-da-ómb-án-íts-á alenje ndí asodzi.
 3-lion 3SM-pst-hit-recip-caus-fv 2-hunter and 2-fisherman
 ‘The lion made the hunters and the fishermen hit each other’.³

In this sentence the Verb Root (VR) **-omb-** ‘hit’ supports the extensions **-an-** for the reciprocal, **-its-** for the causative, and the clitics **u** ‘subject marker’ agreeing in ϕ -features with **mkángo** ‘lion,’ and **da** ‘past tense.’ In brief, the prefixed elements, analyzed as clitics in recent work (Mchombo 2002), includes information identified with Functional Projections within the Principles and Parameters Theory (cf. Chomsky & Lasnik 1993). This paper will focus on argument structure changing morphology, focusing specifically on the issue of argument binding associated with reciprocal constructions (Keenan & Razafimamonjy 2001).

2 Argument Structure and Verbal Suffixation

As indicated, Bantu languages have a number of argument structure changing extensions verbal suffixes that include, *inter alia*, the causative, applicative, stative (or neuter), reciprocal, passive. They also realize the less pervasive and no longer productive reversive, contactive and, positional, morphemes. The verb root (or radical), extended by the extensions, and terminated by the final vowel [a], constitute the Verb Stem (VS). The linguistic significance of the VS is indicated by its being the locus or domain for a number of linguistic processes whose influence does not extend to the clitics. For instance, there is vowel harmony in Chichewa (Mtenje 1985),

² According to Malcolm Guthrie’s (1967-71) classification of Bantu languages, Chichewa is placed in zone N in the unit N31. It is regarded as a dialect variation of Nyanja classified as belong to N30 (see Mchombo 2005, Watkins 1937)

³ The following abbreviations will be adopted:

appl	applicative
assoc	associative marker
ben	benefactive
caus	causative
fv	final vowel
hab	habitual
OM	object marker
pass	passive
pres	present
pst	past
recip	reciprocal
reflex	reflexive
SM	subject marker

Luganda (Katamba 1984), in other Bantu languages, whose domain is exclusively the VS. In Chichewa, verbal reduplication, like vowel harmony, is confined to the verb stem. Below are some constructions illustrating the functioning of the extensions:

- (2) Mkángo u-na-thyól-á mpánda
 3-lion 3SM-pst-break-fv 3-fence
 ‘The lion broke the fence’
- (3) Mkángo unathyóléts-á mbidzí mpánda
 3-lion 3SM-pst-break-caus-fv 10-zebra 3-fence
 ‘The lion made the zebras break the fence’
- (4) Mkángo u-na-thyól-él-á mbidzí mpánda
 3-lion 3SM-pst-break-appl-fv 10-zebra 3-fence
 ‘The lion broke the fence for the zebras’
- (5) Mkango u-na-thyól-éts-él-á mbidzí
 3-lion 3SM-pst-break-caus-appl-fv 10-zebra
 mpánda kwá alenje
 3-fence by 2-hunter
 ‘The lion made the hunters break the fence for the zebras.’
- (6) Mbîdzi zi-na-thyól-éts-él-édw-á mpánda kwá
 10-zebra 10SM-pst-break-caus-appl-pass-fv 3-fence by
 alenje (ndí mkángo)
 2-hunter (by 3-lion)
 ‘‘The zebras got the fence broken for (them) by the hunters at the instigation of the lion’’
- (7) Mbîdzi zi-na-thyól-éts-él-an-á mpánda kwá
 10-zebra 10SM-pst-break-caus-appl-recv-fv 3-fence by
 alenje
 2-hunter
 ‘‘The zebras got the fence broken for each other by the hunters’’

The sentences illustrate the type of arguments affected by the presence of the individual extensions, as well as permissible combinations of those extensions. To the verb **thyola** ‘break’ are suffixed extensions to form the causative **thyoletsa** ‘make break,’ the applicative **thyoleta** ‘break for,’ the applicative of a causative **thyoletsela** ‘get something broken for someone,’

and the passive of the applicativized causative **thyoletseledwa** ‘have something broken for one by’ and the reciprocal of the applicativized causative **thyoletselana** ‘have something broken for each other.’ Morpheme order within the verb stem appears to be constrained by principles that have yet to be fully determined. Proposals concerning morphotactic constraints have ranged from ideas inspired by the “Mirror Principle” advanced by Mark Baker (1985, 1988), through suggestions about morphological template to suggestions that perhaps thematic information is implicated in the linearization of the extensions (for relevant discussion, see Hyman 1991, 2003; Hyman and Mchombo 1992; Mchombo 2004; Ngunga 1997; Sibanda 2004).

With regard to the Mirror Principle the suggestion is that “morphological derivations must directly reflect syntactic derivations (and vice versa)” (Baker 1985:375). This approach ties morphological structure to syntactic derivation, probably determining semantic scope too. Naturally, this gets undermined by morphological orderings that patently conflict with syntactic derivation. Such mismatches between syntactic derivation and morphological order have been noted in some languages, such as Xhosa (cf. Dlayedwa 2002).

The idea of a morphological template that fixes the preferred order of the morphemes on the basis of principles independent of syntactic derivation or semantic composition can be gleaned from the work of Hyman (1991, 2003). Based on studies of various African languages Hyman has noted a recurrence of the order Causative, Applicative, Reciprocal, Passive (CARP). The suggestion is that in the absence of over-riding factors, this is the generally preferred order of those morphemes. Certainly, Chichewa offers some evidence of that in that the applicative and the reciprocal appear in that order even when the syntactic derivation or semantic interpretation would demand reverse order (cf. Mchombo 2004). Discussion of morphotactic constraints will be deferred to other studies.

The relative order of the extensions and the clitics is significant. In Lexical Phonology and Morphology-theoretic terms, the extensions can be analyzed as Level 1 affixes. They are intimately connected to the host VR, deriving the VS. Besides the role that the VS plays in constituting the locus of vowel harmony, reduplication, etc., it is also the input to nominalizations that exclude the prefixed material, as shown below:

kónda ‘love’ / kóndána ‘love each other’ / chikondano ‘mutual love’
 onetsela ‘demonstrate’ / chionetselo ‘exhibition’
 kodza ‘urinate’ / kodzela ‘urinate with’ / chikodzelo ‘bladder’

ongola ‘straighten’ / ongolela ‘straighten with’ / chiongolelo ‘steering wheel’
 tenga ‘take’ / tengana ‘take each other’ / mténgáno ‘death pact’
 ononga ‘damage’ / onongéka ‘get damaged’ / chionongeko ‘destruction’
 senda ‘skin’ / sendédwa ‘be skinned’ / kasendedwe ‘manner of skinning’
 da ‘hate’ / dana ‘hate each other’ / mdani ‘enemy’
 fa ‘die’ / fela ‘die for’ / felana ‘die for each other’ / mafelano ‘intense struggle’

In previous work the following structural representation of the verb in Chichewa has been proposed:

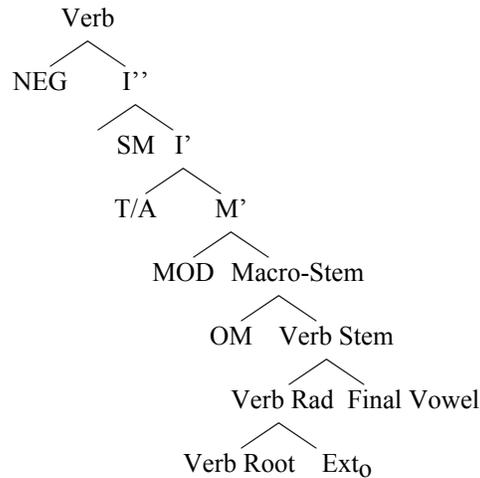


Figure 1

3 On Argument Binding and Syntactic Binding in Chichewa

The concept of binding has played a crucial role in syntactic theory, as evidenced by its inclusion in the name of one major theoretical framework, the theory of Government and Binding (GB) (Chomsky 1981). Binding has traditionally been invoked to account for the construal of two constituents ...X...[...Y...]... within a syntactic configuration such that one is con-

strued as referentially dependent on the other. The dependent one, the bindee, has its construal determined by its binder, the antecedent, under specific structural conditions. The principles of the binding theory specify the elements whose construal is determined by some antecedent, and the requisite structural conditions for successful construal. Binding has, traditionally, involved reflexives and pronominals. The former have their construal with an antecedent resolved within an appropriately defined local domain. The latter, on the other hand, appear to be more involved in discourse structure. Earlier versions of binding theory included reciprocals in the typology of bound elements, grouped together with reflexives as requiring an antecedent within some local domain.

The conflation of reflexives and reciprocals as constituents that are bound under specific conditions of locality is one that seemed to be influenced by the form and distribution of reflexives and reciprocals in English. Thus, given a reflexive such as *'herself'* and the reciprocal *'each other,'* both of which appear in arguments positions in the sentence, and require that their construal be determined within an appropriately defined local domain, it was easy to subsume them under the same category, as anaphors. The sentences below provide the relevant examples:

- (8) a. The queen bought herself a new castle
 b. The baboon and the hyena bought each other presents.

That the anaphors must have their construal determined within the local domain is shown by the ungrammaticality of the following sentences:

- (9) a. *The queen believes that the king will buy herself a new castle.
 b. *The baboon and the hyena believe that the lion will buy each other presents

The identification of the reflexive with the reciprocal as constituting the class of anaphors is one that continues to undergo revision in light of their grammaticalization and semantic properties. In Chichewa, the reflexive is an invariant morpheme *-dzi-* (or its cognates in other Bantu languages), that appears in the position of the object marker (OM). The OM occurs as a left sister of the VS. The OM and the VS comprise the Macrostem (cf. Goldsmith & Sabimana 1985). In Bantu languages the OM has been analyzed as an incorporated pronominal argument anaphorically bound by an antecedent within the discourse structure (cf. Bresnan & Mchombo 1986, 1987; Chimbutane 2003; Deen 2004; Dlayedwa 2002; Letsholo 2003; Rubanza 1988). The grammaticalization of the reflexive, appearing in the

position of an incorporated pronominal argument, argues for its treatment as a pronominal argument whose construal is determined by principles of syntactic binding⁴. Consider the following:

- (10) a. Anyaní a-ku-dzí-mángílil-á ku nthámbí
 2-baboon 2SM-pres-reflex-tether-fv 17-loc 10-branch
 “The baboons are tethering themselves to the branches.”
 b. Anyaní a-ku-wá-mángílil-á ku nthámbí
 2-baboon 2SM-pres-reflex-tether-fv 17-loc 10-branch
 “The baboons are tethering them to the branches.”

The relevant antecedent appears to be the Subject Marker (SM), itself functionally ambiguous as an agreement marker and as an incorporated pronominal argument (cf. Bresnan & Mchombo 1986, 1987; Mchombo 2004). The pronominal argument status of the SM underlies the apparent ‘long distance’ relationship between the reflexive and the constituent that determines its referential value. Take the following:

- (11) Mikángó sí-i-ku-fún-á kutí nkhandwe
 4-lion Neg-4SM-pres-want-fv that 10-fox
 zi-uz-é anyaní kutí í-ma-dzi kânda
 10SM-tell-subjun2-baboon that 4SM-hab-reflex-scratch
 “The lions do not want the foxes to tell the baboons that they (lions) scratch themselves”

The reflexive in the verb *i-ma-dzi-kânda* is anaphorically bound to the NP *mikángo* ‘lions’ in an apparently long-distance anaphoric relationship through the intervention of the SM ‘i,’ an incorporated pronominal acting as the antecedent of the reflexive. In turn, the SM is bound by the NP *mikángo* ‘lions.’ The binding of the reflexive is, certainly, an aspect of syntactic binding in the standard sense, constrained by principles of bound anaphora.

4 The Reciprocal

The reciprocal, on the other hand, is realized as a verbal suffix in Bantu languages. In its morphological realization, the reciprocal is encoded by a

⁴ Sozinho Matsinhe’s account of verbal affixes in the southern African Bantu language of Tsonga departs from treatment of the reflexive as a pronominal argument subject to the principles of syntactic binding. He opts for the analysis of reflexivization as a morpho-lexical process, affecting argument structure. The analysis will be commented upon later.

verbal suffix *-an-*. In some languages the reciprocal is realized by more than one morpheme. For instance, in Luganda and Ci-Yao the verbal suffix **-agan-** is used; in Kikongo, the dialect of Zoombo region of Northern Angola the morpheme **-azyan-** is used (see Carter & Makoondekwa, 1987), and in Runyambo the reciprocal is realized by the morpheme **-angan-** (see Rugemalira 1993). In all these languages the suffix **-an-** remains, sometimes restricted to ‘frozen’ forms whose roots are no longer attestable as independent verbs within the language. In Runyambo, the following reciprocals, derived with **-an-**, lack independently existing verb stems: **bag-an-a** ‘share, divide up’; **bug-an-a** ‘meet’; **fuka-an-a** ‘wrestle, struggle’; **hak-an-a** ‘argue’; **iw-an-a** ‘fight’; **ing-an-a** ‘be equal.’ Reciprocals with the form **-an-** are also derivable from independently attested verb roots. These include forms such as **tong-a** ‘demand payment’ **tong-an-a** ‘quarrel’; **tond-a** ‘create’ **tond-an-a** ‘discriminate against (by origin)’; **nyw-a** ‘drink’ **nyw-an-a** ‘become friends’; **jend-a** ‘go, walk’ **jend-an-a** ‘go together.’

The regular formation of the reciprocal in Runyambo as, *mutatis mutandis*, in the other languages listed above, is through the suffixation of **-angan-** which “...can be attached to most transitive verb roots, provided the derivation makes sense.” ([Rugemalira, 1993 :150). This is shown in such derivations as **nob-angan-a** ‘hate each other’; **jun-angan-a** ‘help each other.’ Some verbs roots allow the suffixation of both **-an-** and **-angan-**, but with different readings. Consider the following data, again from Runyambo:

(12) -reeb-a	look
-reeb-an-a	keep in touch
-reeb-angan-a	look at each other
-kwat-a	hold/touch/seize
-kwat-an-a	be related; stick together;
-kwat-angan-a	hold/seize each other
-ras-a	throw, shoot.
-ras-an-a	fight, struggle.
-ras-angan-a	shoot or throw at each other

In Kikongo the productive reciprocal morpheme is the verbal suffix **-azyan-**, but it also has the suffix **-aan-**. In this language “for some verbs the form is simply **-aan-**, as in **-waanaana** ‘find each other=meet together, and **-monaana** ‘see each other’, but for many it is **-azyaan-**; Thus, one gets **-zola** ‘love’, **-zolazyana** ‘love each other’. Long vowels before NC [nasal

cluster-SAM] are contracted when the addition of the extension brings the number of stem syllables up to four: **-kaamba** ‘give news to’, **-kambazyana** ‘exchange news with each other’, and **-soonga** ‘show’---> **-songazyana** ‘show each other’.” (Carter & Makoondewa, 1987 : 130)

Syntactic configurations in which the reciprocal form appears are largely similar in that the reciprocal requires a plural or group-denoting subject NP. When the group-denoting NP is a coordinate structure, sometimes a comitative construction is used, as shown below, from Swahili:

- (13) Kiboko a-li-vut-an-a na simba
 7-hippo 1SM-pst-pull-recip-fv with 1-lion
 “The hippo and the lion pulled each other”
 Lit. “The hippo pulled each other with the lion”

Such comitative constructions are routinely exploited to overcome various syntactic problems. For instance, the SM is obligatory in verbal morphology, thereby susceptible to analysis as a *de facto* grammatical subject (cf. Demuth & Johnson 1989; Marten & Kempson, 2006). The SM agrees in \square -features with the group denoting NP, a grammaticized TOPIC element (cf. Morimoto 2000). Given the Bantu noun classification system, where the nouns are placed into various gender classes, coordinate structures provide instances where the coordinated nouns may come from different gender classes, resulting in problems relating to the realization of the SM which, somehow, must resolve the gender conflict (cf. Corbett & Mtenje 1987; Mchombo & Ngalande 1980; Mchombo & Ngunga 1994). In such cases, all but the initial conjunct of the participants involved in the action denoted by the reciprocal would be encoded in an extraposed ‘*na NP*’ in Swahili. In some languages, e.g., Ci-Yao, the problem may go beyond issues of resolution of gender conflict. Consider the following sentence:

- (14) a. Coomé cí-kú-ci-súumisy-a nyama cí-súvi
 7-cat 7SM-pres-7OM-sell-fv 9-meat 7-leopard
 “The cat is selling (to) the leopard some meat.”
 b. *Coomé ní cí-súvi yí-kú-súumisy-an-á nyama
 7-cat and 7-leopard 8SM-pres-sell-recip-fv 9-meat

The coordinated nouns in 18b are from the same gender class, and each one of them takes the SM ‘*ci.*’ Their plural forms are *yoóme* ‘cats’ and *yisuví* ‘leopards’ respectively. These belong to class 8, and take the plural SM ‘*yi.*’ The coordinate structure cannot be antecedent to the plural subject marker because the NP does not denote a plurality of cats or leopards. It cannot

bind 'ci' either, the SM for class 7, because that goes with a singular noun whereas the coordinate NP denotes a plural group. Effectively the comitative construction resolves the problems occasioned by such constructions.

5 Argument Binding and Reciprocal Morphology

In Bantu languages, the reciprocal appears to be involved in morpho-lexical operation of verb derivation. The reciprocal derives a one-place predicate from a two-place predicate or, in general, reduces by one the array of arguments associated with the non-reciprocalized predicate. It is a de-transitivizing morpheme that derives predicates with a reciprocal interpretation. Consider the following:

- (15) a. Alenje á-ma-gul-il-á asodzi mikóndo
 2-hunter 2SM-hab-buy-appl-recip-fv 4-spear
 "The hunters buy spears for the fishermen."
 b. Alenje ndí asodzi á-ma-gul-il-án-á mikóndo
 2-hunter conj 2-fishermen 2SM-hab-buy-appl-recip 4-spear
 "The hunters and the fishermen buy each other spears."

In Chichewa the reciprocal participates in all linguistic processes associated with the verb stem. These include reduplication, deverbal nominalization, vowel harmony, etc. Notable is the fact that the reflexive, together with other proclitics, do not participate in either nominalizations of the type indicated, or in reduplication. The deverbal nominals, involving the reflexive, are ungrammatical, as shown below:

- (16) Ku-*dzi*-kónd-a *chi-*dzi*-kondo 'self-love'
 Inf-reflex-love-fv
 'To love oneself'

 Ku-*dzi*-yámík-a *ma-*dzi*-yamiko 'self-praise'
 Inf-reflex-praise-fv
 'To praise oneself'

- (17) mikango í-ma-*dzi*-kánd-a
 4-lion 4SM-hab-reflex-cratch-fv

 *mikango í-ma-*dzi*-kánd-a-*dzi*-kánd-a
 4-lion 4SM-hab-reflex-scratch-fv-reflex-scratch-fv

The reciprocal in Bantu, unlike the reflexive, is not a nominal argument subject to principles of binding. It is a morpho-lexical process deriving reciprocal predicates apparently not susceptible to syntactic binding. However, this is a view that turns out to be controversial.

In a study of Malagasy, an Austronesian language, Keenan and Razafimamonjy observe that the reciprocal is derived morphologically through the affixation of the reciprocal morpheme *-if-* to a transitive verb. The reciprocal *-if-* "...immediately precedes the active prefix *aN-*. Its presence excludes an overt accusative NP..."(Keenan & Razafimamonjy 2001: 41). In this regard the Malagasy reciprocal behaves in a manner comparable to that in Chichewa. However, Keenan and Razafimamonjy state that the position excluded by the presence of the reciprocal remains syntactically active, indicated by the empty category symbol 'e.' They then claim that "[It] is this empty position which corresponds to the presence of the reciprocal pronoun each other /one another in English and which determines one of the arguments, henceforth the reciprocalized argument, of the reciprocal relation used in semantic interpretation of Ss built from reciprocal verbs." (ibid.)

Keenan and Razafimamonjy make the further claim that "...the NP which *-if-* requires to be plural is the *antecedent* of *-if-*. We also say that this antecedent *reciprocally binds* the reciprocal empty category (ec) licensed by *-if-*. The motivation for this terminology is that given an occurrence of *-if-*, the positions determined by its antecedent and its ec are those which determine the arguments of the reciprocal relation used to interpret the reciprocal expression." (ibid. 42).

Having adopted this stance, Keenan and Razafimamonjy pose a number of questions about the antecedent of the reciprocal in Malagasy. These relate to whether the antecedent of the reciprocal

- (a) always occurs external to the VP projected by the verb that the reciprocal morpheme *-if-* occurs in;
- (b) c-commands the reciprocal empty category;
- (c) occurs as an argument of the reciprocal predicate, within the same complete functional complex as the reciprocal verb; and,
- (d) locally binds the empty category licensed by the reciprocal morpheme.

Their conclusion is that for Malagasy answers to questions (a) (b) and (d) are negative. For (c) the answer is affirmative, that the agent phrases of the non-active verbs are arguments of the verb.

The facts about Malagasy are, in relevant respects, comparable to Chichewa. Note, however, that there is a measure of quibbling in the K& R account. On the one hand, the plural antecedent "reciprocally binds" the reciprocal empty category licensed by the reciprocal morpheme. On the

other hand the antecedent does not either c-command or locally bind the empty category licensed by the reciprocal morpheme, as indicated by the negative answers to the questions (b) and (d) above. What then is to be made of the question whether any kind of binding is evident in reciprocal constructions?

The idea of reciprocal binding appears to be rooted in efforts to provide a coherent account of the reciprocal relation that is evident in the semantic interpretation of expressions built from reciprocal verbs. Consider the sentence below:

- (18) a. Alenje ndí asodzi á-ma-lemekez-ân-a
 2-hunter conj 2-fisherman 2SM-hab-respect-recip-fv
 “The hunters and the fishermen respect each other.”
- b. Alenje ndí asodzi á-ma-gul-il-ân-á
 2-hunter conj 2-fisherman 2SM-hab-buy-appl-recip-fv
 mikóndo
 4-spear
 “The hunters and the fishermen buy each other spears.”

The sentences have as their primary readings the claims that the hunters respect the fishermen and the fishermen respect the hunters, and that the hunters buy spears for the fishermen and the fishermen buy spears for the hunters, respectively. Equally, the sentences convey the readings that the hunters respect each other and the fishermen respect each other for 18(a), or that the hunters buy each other spears and so do the fishermen, for 18(b). The latter readings could be derived from sentence coordination, so that will be set aside. Taking the standard approach to accounting for the initial interpretations, the assignment of semantic roles is done in the argument structure. A general constraint in the assignment of semantic roles is that each semantic role be assigned to a particular argument of the predicate and each argument be assigned a single semantic role. Naturally, if reciprocal verbs are derived lexically, and the reciprocal predicate is de-transitivized, then the single argument, the antecedent, must be assigned a single role. Yet, as the semantic interpretation indicates, the group denoted by the antecedent argument requires that the individuals denoted participate both as agents and patients or beneficiaries. How can this be achieved?

The simple clue lies, once again, in the treatment of reflexives that, apparently, resolve the apparent violation of this general constraint on semantic role assignment. A single entity participates as both agent and patient or beneficiary. The reflexive, as a pronominal argument, receives the semantic role of patient. However, as an anaphor, it is bound to the antecedent, effec-

tively resolving the reading that the antecedent is construed as both agent and patient. In other words, in the discourse structure representation the two syntactic arguments are mapped onto the same entity (cf. Sells, Zaenen & Zec 1986). It is but a simple step to the speculation that reciprocal constructions must equally involve binding.

6 On Reciprocal Interpretation

The relevance of the concept of binding to interpretation of sentences with reciprocal predicates is occasionally enhanced by putative analyses of the reflexive in some Bantu languages as a morpho-lexical process, deriving reflexive verbs. In such analyses reflexivization is claimed to be an argument structure reducing process, comparable to the reciprocal. Matsinhe's analysis of the reflexive in Tsonga, a language spoken in Mozambique and South Africa, adopts such a view, exploiting ideas attributed to Jane Grimshaw (cf. Matsinhe 1994). What then is the rationale for keeping them separate?

Matsinhe treats the reciprocal and the reflexive in Tsonga as having comparable effects. The reciprocal morpheme is *-an-*, like in Chichewa and in Bantu in general. He notes that "...this affix changes the predicate argument structure of the verb to which it is attached by binding the object (theme) to the subject (agent), creating co-referentiality. This fact makes the reciprocal affix *-an-* resemble the reflexive prefix *-ti-*. Thus, the former will be treated on par with the latter." (Matsinhe 1994:169). The parity of treatment of the reflexive and the reciprocal is motivated by the observation that "[L]ike the reciprocal affix *-an-*, the reflexive prefix *-ti-* gives rise to coreferentiality between the agent and the theme. The theta role linked to the object is suppressed (bound to the subject), and, as a result, the number of the arguments is reduced by one." (ibid. 170). Matsinhe adopts a suggestion by Grimshaw (1982) that "...reflexivization should be regarded as a morpholexical operation which applies a reflexive lexical rule to the predicate argument structure of a verb, and whose effect is to bind one argument to another" (Grimshaw 1982:106). On that basis, Matsinhe claims that reflexivization affects transitivity patterns, and that "...given a transitive two-place predicate, a reflexive predicate can be derived from it by binding its object to the subject. Hence reflexivization can be regarded as a process which transforms a transitive verb into an intransitive one." (ibid. 170).

Despite its apparent plausibility this analysis of the reflexive is dependent on shifting conceptions of the notion of binding. In general, anaphoric binding deals with the resolution of referential dependencies of pronominal elements. In many Bantu languages, the OMs are incorporated pronominal arguments in anaphorically bound by antecedents outside the minimal clau-

se. In Gikuyu, the OM is in complementary distribution with the object NP (cf. Begvall 1985; Mugane 1997), yet, it is not claimed that the OM detransitivizes the verb in order to account for the omission of the object NP. The grammaticalization of the reflexive which, unlike the reciprocal, appears in the OM position, is somehow discounted as irrelevant to the determination of its status as a pronominal argument that is bound to an antecedent within the clause. Instead, the failure of the verb to support an object NP, comparable to the situation when the OM is present anyway, is construed as evidence that the reflexive is a detransitivizing affix. Note that in languages such as Chichewa which, unlike Gikuyu, allows for clitic doubling, i.e., for the putative object NP to co-occur with the OM, the said object NP is an external Topic (cf. Morimoto 2000).

The reciprocal is associated with the notion of argument binding. Clearly, this is not equivalent to that involved in reflexivization. It has been customary to include reciprocals in discussions of ‘symmetric predicates’ (cf. Lakoff and Peters 1966; McNally 1993). A two-place predicate R , is said to be symmetric if for any two x and y , appropriate arguments of R , the following holds: Rxy is equivalent to Ryx . In other words, if ‘ x is in the relation R to y , but y is not in the relation R to x ’ is contradictory, then R is said to be symmetric. This characterization of symmetric predicates focuses on the intrinsic properties of the relation itself. The situation gets complex when cardinality of the individuals increases and lexical aspects of the predicate are taken into consideration. In some cases, the reading yielded is simply that of group activity. Thus, in the situation of a bar-room brawl, with a large number of individuals, the statement that the people threw bottles at each other does not mean that the relation of ‘throwing bottles’ holds of every pair-wise combination of the people there. Consider, further, the interpretation given to the statement that animals followed each other to the river, or that the substitute players in a sporting event are sitting next to each other on a bench. The interpretations are not comparable to that of say, two individuals shouting insults at each other. The notion of binding as applied to the reciprocal is thus different from that of the reflexive (cf. Alsina 1993; Dalrymple et.al., 1994, 1998; Mchombo 1999b; 2002a, b).

What then is to be made of the idea of argument binding that is associated with the reciprocal? Heim, Lasnik and May (1991) (henceforth HLM) propose to derive the semantics of reciprocals from the morphosyntactic representation for the reciprocal in English. Analyzing *each other* in English as an NP anaphor, the claim, as noted by Keenan and Razafimamonjy, is that “... the patently compound form *each other* undergoes a kind of semantic mitosis into *each* and *other* (at some level) the former interpreted as distributive universal quantifier and the latter as a disjoint reference opera-

tor” (Keenan and Razafimamonjy 2001:80) The analysis of HLM has the reciprocal as having quantificational force, hence susceptible to quantifier-raising (QR) at logical form. The semantics of the reciprocal, involving reciprocal binding can be in part determined from the logical form through scope relations.

Accepting that the reciprocal is quantificational suggests that at LF it should be in an operator position to bind its variable(s). The reciprocal thus introduces quantificational morphology in the Bantu verb stem. Quantifier-raising is a syntactic rule which adjoins a quantifier to its mother node for it to c-command its variable(s) within its scope. The concept of *scope* is itself explicated through the notion of c-command in that the scope of α is the set of nodes that α c-commands at LF. A quantifier’s scope coincides with its c-command domain.

The problem with the reciprocal in Bantu is that as an aspect of quantificational morphology it should be susceptible to QR to achieve the scope effects, but as a constituent of a lexically derived verb stem, it should not be the target of syntactic movement, which would violate lexical integrity. Applying QR to the reciprocal would create a trace in the verb stem, in effect, a variable. However, Baker (1988) notes there are no traces inside words. Clearly, the quantificational morphology associated with the reciprocal in Bantu poses problems regarding the derivation of logical form from the morphological form. Perhaps the reciprocal is offering evidence for not deriving logical form from morphosyntactic organization through movement of constituents. Logical form is linked to, but modeled independently of, morphosyntactic organization. Grammatical theory needs to provide for factorization of natural language into such informational structures as discourse structure, argument structure, functional structure, constituent structure, logical structure, etc., each with its primitives and constraints, providing explicit procedures for capturing the relations among them. Without going into explicit details, this is the architecture that is provided by the theory of LFG (cf. Bresnan 2001; Dalrymple 2001; Falk, 2001).

7 References

- Alsina, A. (1992). On the Argument Structure of Causatives. *Linguistic Inquiry* 23(4): 517-555.
- Alsina, A. (1993). Predicate Composition: A Theory of Syntactic Function Alternations, Doctoral dissertation, Stanford University.
- Baker, Mark (1985) The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16.3: 373-415.
- Baker, Mark. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- Bresnan, Joan (2001). *Lexical Functional Syntax*. Oxford: Blackwell Publishers
- Bresnan, Joan. and Sam A. Mchombo (1987). Topic, Pronoun and Agreement in Chichewa. *Language* 63(4): 741-782.
- Bresnan, Joan and Sam A. Mchombo (1995). The Lexical Integrity Principle: Evidence from Bantu. *Natural Language & Linguistic Theory* 13: 181-254.
- Carter, Hazel and Joao Makoondekwa (1987). *Kongo Language Course: Maloongi Makikoongo. A Course in the Dialect of Zoombo, Northern Angola*. Madison: African Studies Program, University of Wisconsin.
- Chimbutane, Feliciano (2002). Grammatical Functions in Changana: Types, Properties and Function Alternations. M.Phil dissertation, Australian National University.
- Chomsky, Noam (1981) *Lecture on Government and Binding*. Dordrecht, Foris Publications.
- Chomsky, Noam and Howard Lasnik (1993) Principles and Parameters Theory. In (eds.) Jacobs, J, A von Stechow, W Sternefeld, and T.Venneman (1993) *Syntax. An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter.
- Corbett, Greville and Al Mtenje (1987) Gender agreement in Chichewa. *Studies in African Linguistics*, 18. 1-38
- Dalrymple, Mary 2001. *Lexical-Functional Grammar (Syntax and Semantics 34)*. New York: Academic Press
- Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Stanley Peters (1998). Reciprocal Expressions and the Concept of Reciprocity. *Linguistics & Philosophy* 21: 159-210
- Dalrymple, Mary, Makoto Kanazawa, Sam Mchombo & Stanley Peters (1994). What Do Reciprocals Mean? Fourth Semantics and Linguistic Theory Conference: SALT 4, Rochester, NY, Cornell University Working Papers in Linguistics.
- Dalrymple, Mary, Sam Mchombo, and Stanley Peters (1994). Semantic Similarities and Syntactic Contrasts between Chichewa and English Reciprocals. *Linguistic Inquiry* 25(1): 145-163.

- Deen, Kamil Ud 2004. Let's Build Some Agreement: The status of the Subject Prefix in Nairobi Swahili. Paper presented at the 35th Annual Conference of African Linguistics, Harvard University.
- Demuth, Katherine and Mark Johnson (1989). Interaction Between Discourse Functions and Agreement in Setawana. *Journal of African Languages and Linguistics 11*: 21-35.
- Dlavedwa, Ntombizodwa, C. (2002). Valency-Reducing processes in Xhosa. Doctoral dissertation, University of Essex, UK.
- Du Plessis, J. A. and M. Visser (1992). *Xhosa Syntax*. Pretoria, Via Afrika.
- Falk, Yehuda N. (2001). *Lexical Functional Grammar. An Introduction to Parallel Constraint-Based Syntax*. Stanford, CA: CSLI Publications.
- Goldsmith, John & Firmard Sabimana (1985). The Kirundi Verb. Unpublished paper, The University of Chicago.
- Guthrie, Malcolm. (1962). The Status of Radical Extensions in Bantu Languages. *Journal of African Languages 1*: 202-220.
- Guthrie, Malcolm (1967-71). *Comparative Bantu*. Farnborough: Gregg International Publishers.
- Heim, Irene, Howard Lasnik, and Robert May. (1991). Reciprocity and Plurality. *Linguistic Inquiry 22*(1): 63-101.
- Hoffman, Mika. (1991) The Syntax of Argument-Structure-Changing Morphology, Doctoral dissertation, Massachusetts Institute of Technology.
- Hyman, Larry (1991). Conceptual Issues in the Comparative Study of the Bantu Verb Stem. (eds.) Mufwene Salikoko S. and Lioba Moshi *Topics in African Linguistics*. Amsterdam/Philadelphia, John Benjamins Publishing Co.: 3-34.
- Hyman, Larry and Sam A. Mchombo (1992). Morphotactic Constraints in the Chichewa Verb Stem. Papers from the Parasession on the Place of Morphology in a Grammar, Berkeley, CA, Berkeley Linguistic Society.
- Keach, Camilla. N. (1995). Subject and Object Markers as Agreement and Pronoun Incorporation in Swahili. (ed.) A. Akinlabi *Theoretical Approaches to African Linguistics 1*. Trenton, NJ, Africa World Press, Inc. 109-116.
- Keenan, Edward L. & Jean Paulin Razafimamonjy (2001) Reciprocals in Malagasy. UCLA Working Papers in Linguistics. Papers in African Linguistics 1. Harold Torrence (ed). UCLA Department of Linguistics Number 6. 42-89.
- Kiparsky, Paul (1982). Lexical Morphology and Phonology. *Linguistics in the Morning Calm*. T. L. S. o. Korea. Seoul, Korea, Hanshin Publishing Company: 3-92.
- Lakoff, George and Stanley Peters (1966). Phrasal Conjunction and Symmetric Predicates. (eds.) D. Reibel and S. Schane *Modern Studies in English: Readings in Transformational Grammar*. Englewood Cliffs, N.J., Prentice-Hall. 113-142.

- Letsholo, Rose. 2002. Syntactic Domains in Ikalanga. Doctoral dissertation, University of Michigan.
- Mabugu, Patricia (2001). Polysemy and the Applicative Verb Construction in Chishona. Doctoral dissertation, University of Edinburgh.
- Matsinhe, Sozinho (1994). The Status of Verbal Affixes in Bantu Languages with Special Reference to Tsonga: Problems and Possibilities. *South African Journal of African Languages* 14. No. 4: 163-176.
- May, R. (1985). *Logical Form: Its Structure and Derivation*. Cambridge, MA: The MIT Press.
- Mchombo, S. A. (1993). "A Formal Analysis of the Stative Construction in Bantu." *Journal of African Languages and Linguistics* 14: 5-28.
- Mchombo, S. A. (1993). On the Binding of the Reflexive and the Reciprocal in Chichewa. In Mchombo, Sam (ed.) 1993. *Theoretical Aspects of Bantu Grammar*. Stanford, CA: Center for the Study of Language and Information: 181-208.
- Mchombo, Sam. (2002). Affixes, Clitics and Bantu morphosyntax. In Amberber, Mengistu & Peter Collins (eds.) *Language Universals and Variation*. Westport, CT: Greenwood Press. 182-213
- Mchombo, Sam. (2004). *The syntax of Chichewa*. Cambridge, UK: Cambridge University Press.
- Mchombo, Sam (2005). Nyanja. *Encyclopedia of Languages and Linguistics*. 2. Volume 8. Oxford, Elsevier. (ed.) Keith Brown pp.744
- Mchombo, S. A. and R. M. Ngalande (1980). Reciprocal Verbs in Chichewa: A Case for Lexical Derivation. *Bulletin of the School of Oriental & African Studies, University of London* 45: 570-575.
- Mchombo, S. A. and A. S. A. Ngunga (1994). The Syntax and Semantics of the Reciprocal Construction in Ciyao. *Linguistic Analysis* 24(1-2)
- McNally, L. (1993). Comitative Coordination: A Case Study of Group Formation. *Natural Language & Linguistic Theory* 11: 347-379.
- Morapedi, Setumile (2006). Information Structure in Setswana: Locative Inversion and Related Constructions. Doctoral dissertation, University of Sussex, UK.
- Morimoto, Yukiko 2000. Discourse Configurability in Bantu Morphosyntax. Doctoral dissertation, Stanford University.
- Mugane, John. 1997. *A Paradigmatic Grammar of Gikuyu*. Stanford, CA: CSLI Publications.
- Mtenje, A. D. (1985). Arguments for an Autosegmental Analysis of Chichewa Vowel Harmony. *Lingua* 66: 21-52
- Rugemalira, Josephat (1993). Runyambo Verbal Extensions and Constraints on Predicate Structure. Doctoral dissertation, University of California, Berkeley

- Sells, Peter, Annie Zaenen, and Draga Zec (1987). Reflexivization Variation: Relations between Syntax, Semantics, and Lexical Structure." In (eds.) M. Iida, S. Wechsler, and D. Zec. *Working Papers in Grammatical Theory and Discourse Structure. Interactions of Morphology, Syntax, and Discourse*. Stanford, CA: CSLI Publications 169-238.
- Sibanda, Galen (2004). Verbal Phonology and Morphology of Ndebele. Doctoral dissertation, University of California, Berkeley.
- Watkins, Mark Hanna (1937) A Grammar of Chichewa. A Bantu Language spoken in British Central Africa. Language Dissertations, 24. Baltimore, Linguistic Society of America.

Malagasy Instrumental Nominalizations*

DIMITRIS NTELITHEOS

1 Introduction

This paper explores the morphosyntactic properties of two different types of instrumental nominalizations in Malagasy (Austronesian). Both nominalizations are derived by attaching the same nominalizing prefix *f-* to a verbal stem that contains the root and some additional voice morphology:

1. a. n.a.hita [f.an.ala.hidy] aho
 PST.AT.see NML.AT.remove.lock 1SG.NOM
 ‘I found a key. (*Lit. instrument used to remove lock with*)’
- b. ny f.an.doah.an-dRabe ny rindrina dia ilay fantsika
 D NML.AT.drill.CT/LNK-Rabe D wall TOP DEM nail
 ‘The (instrument for) Rabe’s drilling the walls is this nail.’

* I would like to thank Randriamihamina Hasina, Rajaonarison Hery, Rasorofodrainibe Many, Razanajatovo Rado, and Felantsoa Lovamanitra, in Madagascar and Ramahatafandry Noro in Los Angeles for providing the data for this paper. Thanks also to Hilda Koopman, Anoop Mahajan, Ed Keenan and the audiences of AFLA XII and TLS9 for comments and suggestions. All mistakes are my own.

It is proposed that the resulting nominalizations contain syntactic structures of different sizes and that the differences in syntactic behavior of the two nominalizations are directly related the size of structure they contain. In other words, the height of merger for the nominalizing prefix determines the morphosyntactic properties of the derived nominalization. The following table summarizes these properties:

Nominal Type	Internal Arguments		Adv.	Adj.	Implied Event	Episodic Reading	Possessor
	DP	Incorp.					
f-AT (1.a)	NO	YES	NO	YES	NO	NO	YES
f-CT (1.b)	YES	NO	YES	NO	YES	NO	YES

Table 1: Properties of Instrumental Nominalizations in Malagasy.

There are a number of assumptions that will be adopted in the paper without much discussion as they are supported by the data provided here and have been established independently in the related literature. I assume that nominalizations are formed in the syntactic component (c.f. Fu et al 2001, Alexiadou 2001, and others), and contain at least a vP/VP-core with a possible number of further functional projections (Borer 1993; Fu et al. 2001; Hazout 1995)). In terms of categorial status at least some nominalizers are (nominal) complementizers that derive a (reduced) headless relative clause whose structure is assumed to involve raising of a null operator to spec-CP (as discussed for example in Kayne (1994)).

$$[{}_{DP} D [{}_{CP} \mathbf{OP} I [NML \dots [tI]_{VPn} [VP_{n-1}]]]]{}^1$$

Finally, nominalizers are assumed to attach at different heights in the clausal structure, resulting in nominalizations with gradient morphosyntactic properties: the higher the projection where the nominalizer merges the more verbal and less nominal properties the nominalization exhibits (Fu et al 2001, Alexiadou 2001, Schueler, 2004).

¹ The following conventions in abbreviating labels in the examples will be used: D, determiner; SG, singular; PL, plural; AT, agent topic focus or actor trigger; TT, theme topic focus or theme trigger; CT, circumstantial topic; NOM, nominative; GEN, genitive; LOC, oblique/prepositional case usually manifested as prefix *an-*; ASP, aspectual marker; PST, past tense; NML, nominalizer.

mood). The nominalization replaces the aspectual/tense prefix with the nominalizer *f*-. Thus the morphological template for the nominalization should be bracketed as follows:

$$[f \text{ [an-/i- [V}_{\text{ROOT}}]]]$$

The following table contains examples of *f*-AT instrumental nominals:

ROOT	GLOSS	AT-VERB	GLOSS	NOMINAL	GLOSS
a. vaky	'split open'	mamaky	'to cut'	famaky	'axe'
b. zaitra	'sewing'	manjaitra	'to sew'	fanjaitra	'needle'
c. hogo	'haircut'	mihogo	'to comb'	fihogo	'comb'
d. rakotra	'cover'	mirakotra	'to cover'	firakotra	'blanket'

Table 2: *f*-AT nominalizations

In terms of semantic properties, the most prominent feature of these nominalizations is that they denote entities in the world that are [-ANIMATE]. In this respect they contrast with *mp*-nominalizations that are exclusively [+ANIMATE], deriving agentive nominalizations in Malagasy that are similar to English *-er* nominalizations. Contrary to English though, where *-er* nominals denote both actors and instruments (see Rappaport & Levin 1992; Van Hout & Roeper 1998), in Malagasy the two types of nominals have distinct morphology.

In the relevant literature *f*-AT instrumentals are traditionally characterized as "lexicalized" (Keenan and Polinsky 1998; Paul 1996), i.e. forms that are not transparent to syntactic operations. Considering productivity, *f*-AT nominals are the least productive nominals in Malagasy (if we exclude a handful of *f*-TT nominals, see Ntelitheos forthcoming). As far as I know, no new forms have been added to the language for a number of years. A search of Richardson (1885) revealed around forty *f*-AT tokens of instrumental nominals that denote some type of tool (see Ntelitheos (forthcoming) for a complete list). To these we can add forms that could be understood as instrumentals in some sense, but are used primarily for naming species of plants or animals and magical charms used to ward off evil spirits, especially in events like bull-fighting. Their limited productivity may be due to the effects of some sort of 'blocking' (Aronoff 1976:43), i.e. the process that restricts the formation of a potentially regular form due to the presence in the language of another synonymous (or near synonymous) form (e.g. English *thief*/**stealer*). *f*-AT nominals completely exhaust the list of tools that are used in basic everyday human activities. For the derivation of more

specialized instruments and to coin new tools speakers use *f*-CT instrumental nominals (see section 3.2).

Finally, *f*-AT instrumentals are homophonous with a number of productive manner nominalizations, and it is usually the contexts that determines their interpretation. Thus, *fanjaitra* can mean both ‘needle’ and ‘way of sewing’ depending on the context. I will not discuss here the use of *f*-ATs as manner nominals (but see Ntelitheos, forthcoming).

3.2 Nominalizer *f*- + CT form (*f*-CT)

The second type of instrumental nominalization under investigation is formed by prefixing the nominalizer *f*- to the CT form of the verb (2.c). The CT verbal form is derived by attaching the AT prefix *an-/i-* to the verbal root and subsequently attaching the CT suffix *-an*, and a tense prefix, which is null for present tense, *n-/no-* for past or *h-/ho-* for future/irrealis. The nominalization replaces the tense prefix with the nominalizer *f*- (see Table 3)². The morphological template for *f*-CT nominalizations therefore would have the following form:

$$[f- [[an-/i- [V_{\text{ROOT}}]]-an]]$$

Table 3 lists some *f*-CT instrumental nominals with the corresponding roots and verbal forms from which they are derived:

ROOT	GLOSS	CT-VERB	NOMINAL	GLOSS
a. ravona*	(level)	andravonana	fandravonana	‘tool for leveling
b. loaka	‘hole’	andoahana	fandoahana	‘tool opening holes’
c. tovo*	(fill)	anovozana	fanozovana	‘bucket’
d. ady	‘fighting’	iadiana	fiadiana	‘weapon’
e. pasoka	‘ironing’	ipasohana	fipasohana	‘iron’

Table 3: *f*-CT Nominalizations

f-CT nominalizations are very productive with hundreds of forms available and new forms introduced in the language. They form homophonous locative, manner and action or abstract nominalizations with the latter being

² A star symbol (*) next to a root means that the root is not found in isolation. In these cases the meaning of the root is given in parenthesis to show its semantic contribution when deriving verbal or nominal stems.

the most productive. We will consider only their instrumental use in this paper (but see Ntelitheos, forthcoming) for a detailed discussion of all *f*-nominals in Malagasy).

4 Properties of Malagasy Instrumental Nominalizations

4.1 Nominalizer *f*- + AT form (*f*-AT)

Given their morphological shape, *f*-AT nominals must contain at least the lower thematic vP-domain (the root domain) and AT voice morphology. Contra Pearson (2001, 2005) it is assumed here that Voice morphology merges directly above vP and attracts the higher verbal argument in its specifier.

It is also assumed that verbal arguments are NPs with no other functional material available. Consequently, they need to be ‘quantitized’ (i.e. receive Number, Definiteness and Case properties) outside the lower thematic domain and presumably the voice domain (see Sportiche 2005). As a result, any VP-internal arguments cannot surface as (case-marked) DPs but only as incorporated³ arguments. This seems to be confirmed by the data:

3. a. n.a.hita [f.an.ala.hidy] aho
 PST.AT.see NML.AT.remove.lock 1SG.NOM
 ‘I found a key. (*Lit. instrument used to remove lock with*)’
- b. * n.a.hita [f.an.ala.ny hidy] aho
 PST.AT.see NML.AT.remove.D lock 1SG.NOM
 ‘I found a (instrument used to remove the lock with) key.’
- c. * n.a.hita [ny f.an.ala.azy] aho
 PST.AT.see D NML.AT.remove.3SG.ACC 1SG.NOM
 ‘I found the instrument one removes it with.’

As we can see, in (3.b-3.c) the *f*-AT nominal cannot contain a definite DP (determiner + NP or pronoun). Therefore the projections where these phrase merge cannot be available in the contained structure.

³ I assume that the incorporated argument is the NP in VP-internal position without any functional layers. This is supported by the fact that these NPs are not referential and cannot be modified by adjectives or relative clauses.

- b. hita.ko [ny f.an.oto roa]
 see.1SG/GEN [D NML.AT.pound two]
 ‘I found the two pestles.’
- c. hita.ko [ny f.an.oto rehetro]
 see.1SG/GEN [D NML.AT.pound all]
 ‘I found all the pestles.’
- d. hita.ko [ny f.an.oto (izay) no.vid.in-dRabe]
 see.1SG/GEN [D NML.AT.pound (that) PST.buy.TT/LNK-Rabe]
 ‘I found the pestle that Rabe bought.’

Therefore, the whole range of nominal modifiers is available in *f*-AT instrumental nominalizations and consequently most of the hierarchical nominal functional domain must be contained in these nominalizations.

4.2 Nominalizer *f*- + CT form (*f*-CT)

Turning now to *f*-CT nominalizations, given their morphological shape and syntactic behavior, it is assumed that they contain a richer syntactic structure including the lower (thematic) domain, voice morphology and the higher functional domain where internal verbal arguments are ‘quantized’. The prediction then is that these nominalizations should allow for independent expression of the internal argument as a (case-marked) DP:

8. [ny f.an.doah.an-dRabe (ny) rindrina] dia ilay fantsika
 D NML.AT.drill.CT/LNK-Rabe (D) wall TOP DEM nail
 ‘The (instrument for) Rabe’s drilling (the) walls is this nail.’

Given the availability of higher functional projections where internal arguments can be quantized, one would expect incorporation not to be an available option for *f*-CT nominals. The data seem to indicate that this prediction is wrong:

9. [ny f.an.doah.an-drindrin-dRabe] dia ilay fantsika
 D NML.AT.drill.CT-wall.LNK-Rabe TOP DEM nail
 ‘Rabe’s (instrument for) drilling walls is this nail.’

However, the NP-theme in (9) can be modified by adjectives or relative clauses, and can be referential. Therefore, I term this process pseudo-incorporation (following Massam 2001) and assume that the NP has vacated the VP and has been ‘quantized’ in some projection XP, outside the

thematic and voice domains. Pseudo-incorporation takes place when the predicate (possibly VoiceP) inverts and lands in a position phonologically adjacent to the NP-theme (in other words no overt material intervenes between them).

Going back to the possibility of interpreting the instrumental nominalization as eventive, we see that contrary to *f*-AT nominals, *f*-CT nominals are interpreted as eventive when a definite argument is present:

10. ny f.an.doah.an-dRabe (ny) rindrina dia ilay fantsika
 D NML.AT.drill.CT/LNK-Rabe (D) wall TOP DEM nail
 ‘The (instrument for) Rabe’s drilling (the) walls is this nail.’

In (10) some event of ‘wall-drilling’ must have taken place.

One thing that distinguishes morphologically *f*- nominals from their clausal counterparts is the lack of tense morphology. In current approaches on the syntax/semantics of tense (c.f. Zagana 1990; Stowell 1996), tense orders the event relative to some reference time. This time is the moment of speaking in main clauses or the main predicate’s event time in subordinate clauses. This partition of the time reference information predicts that only when tense is syntactically present is there obligatory anchoring of the event to some reference time. If tense is not available, an event may be implied but it is not necessarily anchored. The prediction then is that tenseless nominalizations may not acquire an episodic interpretation but can be interpreted only as habitual⁵. This is true for the example in (10). If an episodic reading needs to be forced, then the corresponding tensed headless relative clause is used:

11. ny n.an.doah.an-dRabe rindrina omaly dia ilay fantsika
 D PST.AT.drill.CT/LNK-Rabe wall yesterday TOP DEM nail
 ‘The (instrument for) Rabe’s drilling walls yesterday is this nail.’

f-CT nominalizations exhibit a distribution of adverbial modifiers that is identical to the one exhibited in clauses (compare (12.a) to (12.b)):

12. a. n.an.doah.an-dRabe rindrina tsara ilay fantsika
 PST.AT.drill.CT/LNK-Rabe wall well DEM nail
 ‘This nail, Rabe drilled walls well (with it).’

⁵ This still leaves the option of allowing tenseless nominalizations with an episodic interpretation when some other mechanism can anchor the event relevant to the moment of speaking. In Malagasy, for example, this can be done when a preposition or locative adverbial is present, since both of these elements carry tense marking.

- b. ny f.an.doah.an-dRabe rindrina tsara dia ilay fantsika
 D NML.AT.drill.CT/LNK-Rabe wall well TOP DEM nail
 ‘The (instrument for) Rabe’s drilling walls well is this nail.’

On the other hand adjectival modification is not possible with *f*-CTs that contain a definite theme argument:

13. * ny f.an.doah.an-dRabe rindrina kely dia ilay fantsika
 D NML.AT.drill.CT/LNK-Rabe wall small TOP DEM nail
 ‘The small (instrument for) Rabe’s drilling walls is this nail.’

Finally, *f*-CT nominalizations license possessors:

14. ny f.an.doah.an-drindrina-dRabe dia ny fantsika
 D NML.AT.drill.CT-wall.LNK-Rabe TOP D nail
 ‘Rabe’s (instrument for) drilling walls is a nail.’

Given that in the corresponding verbal clauses the linked element is an actor, it is not immediately clear why *Rabe* in (14) should be interpreted as a possessor. However, as Ntelitheos (2005b) has shown, when pseudo-incorporation takes place, the linked element is a possessor. Some evidence for this comes from examples like (15):

15. ?? ny f.an.doah.an-drindrina-dRabe dia an-dRaso
 D NML.AT.drill.CT-wall.LNK-Rabe TOP LOC-Raso
 ‘Rabe’s (instrument for) drilling walls is Raso’s.’

In (15) the linked element *Rabe* is interpreted as a possessor and therefore a second possessor *Raso* is pragmatically irregular. No problem arises in sentences where the linked element is interpreted as an actor (see Ntelitheos 2005b for discussion on the interpretation of linked subjects in Malagasy):

16. ny f.an.doah.an-dRabe rindrina dia an-dRaso
 D NML.AT.drill.CT/LNK-Rabe wall TOP LOC-Raso
 ‘The instrument for Rabe’s drilling walls belongs to Raso.’

5 Analysis

It is assumed that the nominalizing prefix *f-* is a nominal complementizer (C^0). We understand the label C^0 as a general label for functional elements in accordance with recent proposals on the syntax of functional elements as complementizers (c.f. the analyses of English *of* and French *de* in Kayne 2002, other prepositional elements in Cinque 2006, C/D nominal clauses in Maasai, in Koopman 2005, and others). Given its status as a nominal complementizer, *f-* merges at different levels in the clausal spine creating reduced CP domains that have the structure, semantics, and (sometimes) distribution of relative clauses. I will adopt here, without further discussion, Kayne's (1994) analysis of relative clause structure, where a determiner selects for a clausal string (CP), the specifier of which is the landing site of the raising of what is traditionally called the 'head' of the relative clause. However, I will assume that the domain of the nominalization is something smaller than a DP (see also Koopman 2005 where the functional element is labeled D/C), its size being determined by the height of merger of the nominalizer. Therefore, there may be available space for nominal modifiers to merge between the nominalization and the definite determiner. In this sense the nominalizer acts as a 'linker' (c.f. Den Dikken and Singhapreecha 2004; Den Dikken forthcoming). In Den Dikken's approach a 'linker' is any functional element that licenses a predicate or its subject in its specifier. In the case of Malagasy nominalizations, *f-* is a linker that attracts the subject/null operator to its specifier (see Ntelitheos (forthcoming) for detailed discussion).

Given the data discussed in Section 4.1, in *f*-AT clauses the initial structure seems to contain at least vP, i.e. the domain where the thematic properties are satisfied:

$$17. \quad [{}_{CP} f- [{}_{VOICE} \text{an-} [{}_{VP} [{}_{VP1} \text{ INSTR} [{}_{VP2} \text{ THEME}]]]]]$$

VOICE
THEMATIC
DOMAIN

VoiceP attaches and creates the projection for the highest argument to be licensed. Since in these cases the actor is always unexpressed (or not present as in Alexiadou's (2001) "light" v account) the highest argument is the instrumental operator which is licensed in spec-vP. The fact that both instruments and actors can be licensed in spec-vP accounts for the empirical generalization that in numerous languages agentive and instrumental nominalizations are expressed morphologically in a similar way (c.f. Comrie & Thompson 1985). Subsequently, the nominalizer attaches, creating a reduced relative and attracting the operator to its specifier. The resulting structure is interpreted roughly as '(instrument) (one) VPs with'.

The projection where the theme is quantized is not available and thus the only remaining option for the internal argument is to appear incorporated inside the VP. Adverbs merge above VoiceP and thus no adverbial modification is allowed.

The headless reduced relative has the distribution of an NP. With respect to its extended projection it behaves like common nouns in that it allows for adjectives, numerals, quantifiers, relative clauses and possessors to be licensed (examples (6) and (7.a-7.d); see Ntelitheos 2005a, 2005b, for discussion of possessor licensing and adjectival modification in these nominalizations):

18. $[_{DP} ny \dots [_{PossP} \dots [_{AP} \dots [_{CP} f- [_{VoiceP} an- [_{VP} [_{VP1} INSTR [_{VP2} THEME]]]]]]]]]]$
 NOMINAL DOM. VOICE THEMATIC DOMAIN

Turning now to *f*-CT clauses, we have seen in Section 4.2 that they exhibit a greater number of clausal properties than *f*-AT nominals. In the analysis adopted here this seems to indicate that more verbal functional layers are contained in these nominalizations. The bracketed representation in (19) exhibits the minimum structure (details omitted) that they contain:

19. $[_{DP} [_{CP} [_{AdvP} [_{EventP} [_{DefP} [_{CaseP} [_{XP} [_{Voice(CT)P} [_{VoiceP} [_{VP} [_{VP1} [_{VP2}]]]]]]]]]]]]]]]]]$
 FUNCT. DOMAIN QUANT. VOICE THEMATIC DOMAIN

The structure now is ‘big’ enough to include the projection where themes are quantized and thus definite themes are expected to appear with *f*-CT nominalizations (see example (8)). If the theme is indefinite it stops at spec-XP and the predicate (VoiceP) inverts over it allowing for phonological adjacency between VoiceP and the Theme. Thus, pseudo-incorporation can take place.

The clausal structure contains at least one aspectual projection (EventP), which binds the event variable (Travis 2000a, 2000b; Pearson 2001, 2005). An event co-occurs with definite themes and thus it is assumed to merge above DefP. Since EventP is available inside *f*-CT nominalizations that contain a quantized theme, an event is implied in these cases. This functional domain is also where adverbs merge and thus adverbial modification is allowed (see example (12)). Since the domain where adverbial modifiers merge is contained in the clausal string within the nominalization no adjectival modifiers are expected (13). This is compatible with proposals that assume that the nominalizer interrupts and changes the ‘extended projection’ (Grimshaw 2000) of the lexical base (c.f.

Schueler 2004). Finally, possessors merge outside the nominalization and therefore they are expected to appear with *f*-CT nominalizations (14).

6 Conclusion

I have shown that the distribution of instrumental nominals in Malagasy is explained if we assume that nominalizers can attach at different heights in the clausal structure. The morphological templates of section (3) are shown to be uninformative with respect to the morphosyntactic properties and distribution of these nominalizations. It has been shown that if these templates are translated into the syntactic structures of (17-19) we obtain a much more elegant account of the syntactic behavior of *f*-nominals in Malagasy. The broader consequence is that processes that are traditionally assumed to fall within a separate morphological component can be explained by using independently motivated syntactic tools. A number of theoretical and empirical issues need to be resolved. One of the fundamental questions that this research raises is what restricts the distribution of nominalizers intra- and crosslinguistically and how to account for gaps in their distribution. For example, why it is not possible to form *f*-nominalizations with the TT form in Malagasy (excluding a few unproductive forms)?

References

- Alexiadou, Artemis. 2001. *Functional Structure in Nominals: Nominalization and Ergativity*. Amsterdam: John Benjamins.
- Aronoff, Marc. 1976. *Word Formation in Generative Grammar*. Cambridge, Mass.: MIT Press.
- Borer, Hagit. 1993. *Derived Nominals*, ms., University of Massachusetts, Amherst.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads; A Cross-Linguistic Perspective*. Oxford: Oxford University Press.
- Cinque, Guglielmo. 2006. Complement and Adverbial PPs: Implications for Clause Structure. *Restructuring and Functional Heads. The Cartography of Syntactic Structures*, vol.4, New York and Oxford, Oxford University Press.
- Comrie Bernard and Sandra A. Thompson. 1985. Lexical nominalization. *Language Typology and Syntactic Description Vol. 3*. ed. T. Shopen. 349-398. CUP Cambridge.
- Den Dikken, Marcel. Forthcoming. *Relators and Linkers: The Syntax of Predication, Predicate Inversion and Copulas*. Cambridge, Mass.: MIT Press.
- Den Dikken, Marcel and Pornsiri Singhapreecha. 2004. Complex noun phrases and linkers. *Syntax 7: 1-54*.

- Fu, Jingqi, Thomas Roeper & Hagit Borer. 2001. The VP within Process Nominals: Evidence from Adverbs and the VP Anaphor Do-So. *Natural Language and Linguistic Theory* 19:549-582.
- Grimshaw, Jane. 2000. Locality and Extended Projection. Lexical Specification and Insertion, eds. P. Coopmans, M. Everart, and J. Grimshaw. 115-133. John Benjamins. Amsterdam/Philadelphia.
- Hazout, Ilan. 1995. Action nominalizations and the lexicalist hypothesis. *Natural Language and Linguistic Theory* 13:355-404
- Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass: MIT Press.
- Kayne, Richard S. 2002. On some prepositions that look DP-internal: English of and French de. *Catalan Journal of Linguistics* 1:71-116.
- Keenan, Edward L. & Maria Polinsky. 1998. Malagasy (Austronesian). *The Handbook of Morphology*, eds. A. Spencer & A. Zwicky. 563-623. Oxford: Oxford University Press.
- Koopman, Hilda. 2005. On the parallelism of DPs and clauses: Evidence from Kisongo Maasai. *Verb First*, eds. A. Carnie, H. Harley and S. A. Dooley. 281-301.
- Massam, Diane. 2001. Pseudo- noun-incorporation. *Natural Language and Linguistic Theory* 19:153-197.
- Ntelitheos, Dimitrios. 2005a. Malagasy Participant Nominalizations: A Structural Account. *Proceedings of AFLA XII, UCLA Working Papers in Linguistics 12*, eds. J. Heinz & D. Ntelitheos. 265-281. Department of Linguistics, UCLA.
- Ntelitheos, Dimitrios. 2005b. Variation in the Distribution of Genitive Subjects in Malagasy. 35th Annual Meeting of the Michigan Linguistic Society (MLS 35), Michigan State University, East Lansing, MI, Saturday, Oct. 15, 2005.
- Ntelitheos, Dimitrios. Forthcoming. The Morphosyntax of Nominalizations: A Case Study. PhD Dissertation. UCLA.
- Paul, Ileana. 1996. F-words and other (Malagasy) nominals. *Proceedings of Canadian Linguistics Association, Calgary Working Papers in Linguistics*. 321-332.
- Paul, Ileana. 2000. Malagasy Clause Structure. Ph.D. Dissertation, Department of Linguistics, McGill University.
- Pearson, Matthew. 2001. The Clause Structure of Malagasy: A Minimalist Approach. *UCLA Dissertations in Linguistics 21*, Department of Linguistics, UCLA.
- Pearson, Matthew. 2005. The Malagasy subject/topic as an A' -element. *Natural Language & Linguistic Theory* 23: 381-457.
- Rappaport, H. Malka & Beth Levin. 1992. -Er Nominals: Implications for the Theory of Argument Structure. *Syntax and Semantics 26: Syntax and the Lexicon*, eds. T. A. Stowell & E. Wehrli. 127-153. New York: Academic Press.
- Richardson, James. 1885. *A New Malagasy-English Dictionary*. LMS, Antananarivo. Reprinted in 1967 by Gregg Press, Farnborough, Hants.

- Schachter, Paul. 1987. Tagalog. *The World's Major Languages*, ed. B. Comrie. 936–958. Oxford University Press, New York.
- Schueler, David. 2004. Extended Projection, Categorization, and the English Morpheme *-ing*. Western Conference of Linguistics (WECOL) 2004, University of Southern California, Los Angeles, November 12-14, 2004.
- Sportiche, Dominique. 2005. Division of Labor between Merge and Move: Strict Locality of Selection and Apparent Reconstruction Paradoxes. *LingBuzz*, available at <http://ling.auf.net/lingbuzz/@OIUETTjffmWuiVMq/GnKnPeKC>
- Stowell, Timothy. 1996. The Phrase Structure of Tense. *Phrase Structure and the Lexicon*, eds. J. Rooryck & L. Zaring. Dordrecht: Kluwer.
- Travis, Lisa de Mena. 2000a. The L-syntax/S-syntax boundary: Evidence from Austronesian. *Formal Issues in Austronesian Linguistics*, eds. I. Paul, V. Phillips & L. Travis. 167-194. Dordrecht: Kluwer.
- Travis, Lisa de Mena (2000b) Event Structure in Syntax. *Events as Grammatical Objects*, eds. Tenny, C. and J. Pustejovsky. 145-185. Stanford: CSLI.
- Van Hout, Angeliek and Thomas Roeper (1998) Events and Aspectual Structure in Derivational Morphology. *Papers from the Upenn/MIT Roundtable on Argument Structure and Aspect*, vol. 32, ed. Heidi Harley. 175-200. MIT Papers in Linguistics, Cambridge, Mass.
- Zagona, Karen. 1990. Times as Temporal Argument Structure. MS, Univ. of Washington.

A Reanalysis of Nonemphatic Pronouns in Dagbani

Tristan Michael Purvis, Indiana University, Bloomington

1 Introduction

The clitic-like pronouns in Dagbani, referred to as ‘nonemphatic’ or ‘weak’ pronouns, are described as being marked for position in relation to the verb as opposed to grammatical role (Wilson 1973, Bawa 1980, Yahaya 1995, Olawsky 1999). However, in the course of compiling a broad corpus of written and spoken Dagbani texts,¹ two pieces of evidence overlooked or undocumented by previous linguists (pronoun conjunction and qualified pronouns), combined with one well-known piece of evidence (identical sets of ‘possessive’ and ‘preverbal’ nonemphatic pronouns), prove problematic for this description and point towards yet another account of these pronouns—a primarily phonologically motivated division based on processes of cliticization. Additional linguistic phenomena (such as vowel elision with emphatic pronouns) and theoretical issues (such as debate of the inventory of phonemic vowels in Dagbani) corroborate this analysis and are discussed in this paper. In the following sections, I will describe the previous grammatical descriptions of the system of personal pronouns in Dagbani and

¹ The data used in this analysis is drawn from this corpus—compiled in 2003/2004 on Fulbright-Hays Doctoral Dissertation Research funding—which comprises 356 spoken and written texts from a variety of contexts and contains a total of approximately 180,000 words. Inquiries may be sent to tpurvis@indiana.edu.

Texas Linguistics Society 9: Morphosyntax of Underrepresented Languages

Douglas S. Bigham, Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, and Jessica White (vol. eds.) and Stephen Wechsler (series ed.).

Copyright © 2007, CSLI Publications.

make recommendations for a new analysis based on this overlooked evidence. By way of a discussion on the interface between syntax and phonology in the realization of nonemphatic pronouns, the paper also includes an argument against the claim that Dagbani is a language exemplifying ‘low selectivity’ on the part of clitics.

2 Overview of Dagbani Personal Pronouns

Dagbani, a Gur language (Volta-Congo) spoken by approximately 500,000 speakers in Northern Ghana, is characterized by two sets of personal pronouns: emphatic or ‘strong’ pronouns and nonemphatic or ‘weak’ pronouns. As the terms ‘emphatic’ and ‘nonemphatic’ are used in the most current grammatical description of the Dagbani language, I will begin by employing these terms. However, after presenting a new description of the ‘nonemphatic’ pronouns, I will argue that the terms ‘weak’ and ‘strong’ are the more appropriate. The thesis of this paper mainly concerns the nonemphatic pronouns; however, I will first present a description of the emphatic pronouns, which will later have a bearing on the analysis of the system of nonemphatic personal pronouns.

2.1 Emphatic Pronouns

The category of emphatic or ‘strong’ pronouns is generally accepted as one distinct set, with little dispute or confusion in terms of description of the variant forms. They are presented in Table 1.

	Singular	Plural
1 st person	mani (man’)	tinima (tin’, tinim’)
2 nd person	nyini (nyin’)	yinima (yin’, yinim’)
3 rd person [+animate]	ɲuna (ɲun’)	bana, (ban’, banima, banim’)
3 rd person [–animate]	dina (din’)	ɲana (ɲan’) †

† Cited in Olawsky (1999), but this form is described as growing obsolete and mainly found in formal texts. In my corpus, this form is actually only attested as a relative pronoun (which are drawn directly from the set of emphatic pronouns). Colloquially, *dina* is said to be used also for 3P inanimate. In practicality, however, a demonstrative is just as likely employed versus an emphatic pronoun, and one form of demonstrative happens to be homophonous with the emphatic pronoun (*dina*)—or perhaps the relationship is one of polysemy. The plural of this is formed with the suffix *-nima* (*dinnima*).

Table 1. Dagbani Emphatic Pronouns

Emphatic pronouns reflect only the number, person, and animacy of their referent, but are not marked for grammatical role or any other function. Bawa (1980) does attempt to distinguish the elided forms in Table 1 as ‘Emphatic’ and the non-elided forms as ‘Disjunctive’; however, these are

simply two functions of the same set of pronouns. (Furthermore, the ‘Emphatic’ function can in fact be represented by non-elided forms (see, e.g., (1c) in the data set further below); and the ‘Disjunctive’ pronouns may be pronounced in elided form (see, e.g., (1f)).) The only variation in the realization of these pronouns (noted in parentheses in Table 1) is the muting of the final *-i* or *-a* when they come before another word and truncation of the sequence *-nima* to *-n*’ in the plural forms. While there are presumably limited constraints on the variable use of full versus truncated forms, emphatic pronouns can essentially appear in most any grammatical role and may be conjoined or adjoined freely with most any variety of nouns and pronouns. The data in (1) show a sampling of the various environments in which these pronouns occur.

- (1) a. **man**’ ni wurim ya ŋɔ
 1S/EMPH FUT demolish town DEM
 ‘I will demolish this town.’ (Subject)
- b. **mani**, **man**’ n- me dimbɔŋɔ zaa
 1S/EMPH 1S/EMPH INF/EMPH build DEM all
 ‘Me, [It is] I [who] built all this.’ (Disjoint, Emph. Subj.)
- c. di yi nyɛ **mani** ka bii yi yen chaŋ lala
 3S/INAM-A COND COP 1S/EMPH CONJ child COND about.to go thus
 ‘If it were **me**, and [my] child was about to go like that’ (Post-copula)
- d. ŋun ŋma n- pa **mani**, **man**’ ŋubi o
 3S/REL/ANIM cut INF/SEQ layupon 1S/EMPH 1S/EMPH chew 3S/ANIM
 ‘Whoever deceives **me**, I[’ll] eat him up’ (Object, Subject)
- e. **man**’ suhupielli nyɛ -la so karim li
 1S/EMPH happiness COP FOC INDEF/SING/INAM read 3S/INAM-B
 ‘My great wish is that someone read it.’ (Possessive)
- f. **yinim**’ ka n yeli ŋŋɔ maa
 2P/EMPH FRO 1S-A tell thus DET/TOP
 ‘It’s **you** whom I’m saying this to.’ (Fronted Ind. Obj.)

Emphatic pronouns are most typically used as disjoint pronouns as in a one-word response to a question such as ‘who’s there?’ or as in the disjoint topic pronoun at the beginning of (1b), and as subject pronouns as in (1a), often co-occurring with the emphatic verb construction as with the second emphatic pronoun in (1b). However, they can also be found in object and possessive grammatical positions and after the copular verb as exemplified in (1c-e). If a solitary personal pronoun is ever fronted with the *ka* fronting construction, the emphatic pronoun must be used, as exemplified in (1f).

2.2 Nonemphatic Pronouns

The nonemphatic or ‘weak’ or ‘simple’ pronouns, however, have proven more problematic to describe. Like the emphatic pronouns, they are marked for person, number, and animacy. In addition, however, after accounting for person, number, and animacy, half of the pronouns in this category appear in two forms, and the variation is not simply a case of a muted vowel or truncation as with the emphatic pronouns. Under a cursory analysis, the variation appears to be based on a structural, grammatical consideration akin to the ‘subject/object’ grammatical case distinction in English or many other languages of the world. Indeed, the average educated Dagbani speaker, influenced by English grammar, is likely to describe these in such terms. For example, a bilingual handbook for the formal teaching of Dagbani in schools presents mutually exclusive sets of these pronouns in ‘object’, ‘subject’, and ‘indirect object’ functions (Mohammed 2004).

Among linguists, however, the nonemphatic pronouns in Dagbani are generally described as being marked for position in relation to the verb, as opposed to grammatical case (Bawa 1980, Olawsky 1999, Wilson 1973, Yahaya 1995). Wilson refers to ‘after-verb’ forms versus pronouns ‘preceding nouns and verbs.’ Olawsky (1999) refers to ‘preverbal’ versus ‘postverbal’ forms. Bawa (1980), who mixes terms by including an ‘Objective’ category of pronouns, later refers to these as ‘post-Verb’ or ‘After Verb’ pronouns, following Wilson’s analysis. Yahaya (1995) also speaks of ‘Post-V’ forms, implying that the other forms are default or underlying. Table 2, adapted from Olawsky² (1999), presents the most current formal linguistic account of Dagbani nonemphatic pronouns.

	Singular			Plural		
	Preverbal (A)	Unmarked	Postverbal (B)	Preverbal (A)	Unmarked	Postverbal (B)
1 st person	ŋ [†]		ma	[tɛ]	ti	[ti]
2 nd person		a		yi [yi]		ya
3 rd person [+animate]		o		bɛ [bɛ]		ba
3 rd person [-animate]	di [dɛ]		li	di [dɛ]	(ŋa) ^{††}	li

[†] homorganic nasal (with allomorphs of [n], [m], [ŋ] and ‘allographs’ of *n* and *m*)

² Following most grammatical descriptions, the terms in this table, as well as in the sample data, are presented using standard Dagbani orthography which is not consistently reflective of phonetic pronunciation. Of particular importance, as noted later, is the fact that, with the exception of [yi], the vowels of the preverbal forms are all ‘muted-i’ [ɛ] ([dɛ], [bɛ], and [tɛ]). Phonetic pronunciation will be clarified wherever necessary within this analysis.

^{††} *ya* is described as becoming obsolete. More is said on this in Section 4.3.7.

- Note: in the linguistic data presented in this paper, unless unmarked for position, ‘preverbal’ pronouns are coded as (A) and ‘postverbal’ pronouns are coded as (B).

- The watermarked/bracketed items reflect phonetic pronunciation (see Footnote 2).

Table 2. Dagbani Nonemphatic Pronouns

Incidentally, possessive pronouns pattern exactly the same as the ‘preverbal’ (and unmarked) set of pronouns. Olawsky (ibid.) presents the possessive pronouns as a separate set of pronouns in the Dagbani pronominal system. Wilson (1973) on the other hand treats the ‘preverbal’ pronouns and possessive pronouns as a common set described simply as ‘pronouns preceding nouns and verbs,’ thereby economizing his grammatical account.

This categorization of pronouns in Dagbani is primarily motivated by the observation of what might otherwise appear to be ‘object pronouns’ in the preterit position after the copular/equative verb *nyɛ* as in (2).

(2) a.	ala	n-	nyɛ	li ?	[ala n- nyɛ *di]
	howmuch	INF/EMPH	COP	3S/INAM-B	
	‘How much is it?’				
b.	paɣa	n-	nyɛ	ma	[paɣa n- nyɛ *n]
	woman	INF/EMPH	COP	1S-B	
	‘I am a woman.’				

According to Wilson (ibid.), another motivation of this analysis is that the *ya* form of the 2nd person plural pronoun follows the verb in the plural imperative construction as in *nyamiya!* (‘look!’). However, Dagbani orthography and current views of Dagbani grammar (Olawsky 1999) analyze this as a distinct suffix, not as a pronoun—not to mention the fact that some speakers let the same *-ya* follow the verb with 1st person plural imperative forms as in *ti cheliya gadama!* (‘let’s end gossip!’). Wilson (1973) may be correct in assuming that this construction derives historically from the pronoun *ya*, but this is arguably not a relevant factor in synchronic analysis.³

3 Evidence Contradicting Descriptions Based on Position in Relation to Verb

On closer analysis of naturally occurring data, the patterning of Dagbani nonemphatic pronouns is a little more complex than these previous grammatical sketches have described.

³ In Mampruli, a sister language of Dagbani generally viewed to be more preservative of the proto language, the 2nd person plural imperative is formed by *ya* (ordinary 2P pronoun) plus the imperative form of the verb (Naden, p.c.).

3.1 Pronouns in Conjoined Noun Phrases

When pronouns appear in conjunctive Noun Phrases with the nominal conjunction *mini* or *ni* ('and'), we find deviations from the preverbal/postverbal categorization, as shown in (3).

- (3) a. m mini ba n- pun bala
 1S-A CONJ 3P-B INF/EMPH now 'there are'
 'It is they and I that are now there.' (**They and I** are now there.)
- b. n mini ba kana
 1S-A CONJ 3P-B come
 'They and I have come.' / 'I have come with them.'
- c. yi bɔli n mini Warahima
 2P call 1S-A CONJ Warahima
 '[... that] you call me and Warahima.'

The categorization of pronouns based on position in relation to the verb would have predicted such forms as *m mini *be kana* and *yi bɔli *ma mini Warahima*. However, in all relevant data, no such forms are attested and are presumed to be ungrammatical. Table 3 presents a tally of all instances of variable nonemphatic pronouns in conjoined NPs.

	PRO-A (i.e., 'preverbal')	PRO-B (i.e., 'postverbal')
X <i>mini</i> + PRO + Verb	0	17 (1 ya, 16 ba) [†] Plus 7 in written (6 ba, 1 li)
Verb + PRO + <i>mini</i> X	5 (3 n, 1 be, 1 di) Plus 1 fronted object 'n mini X' Plus 3 in written (1 n, 2 di)	0

[†] Examples with 1S *ma* are not attested because of supposed ordering restrictions of pronouns in conjoined NPs (Bawa 1980: 85). However, Bawa mistakenly includes an example that contradicts this hierarchy ('*A mini ma ni bana*'). Of interest here is the fact that his contradictory, hypothetical form uses *ma*, not *n*.

Table 3. Nonemphatic Pronouns in Conjoined NPs (targeted cases)

Unless otherwise noted, all example sentences presented in this paper are from spoken texts in my corpus. Notoriously fraught with typographical errors and inconsistency in orthographical conventions, and having an orthography that does not consistently reflect phonology (Olawsky 1999); written data would not be a reliable source upon which to base an analysis of Dagbani pronouns, especially an analysis that sees this as a primarily phonological phenomenon. However, even the written data tend to reinforce the findings based on spoken utterances; so, results from the written corpus are also provided separately.

Wilson (1973) also apparently observed forms such as these in his research, as he comments, at the end of his brief explanation of why ‘objective’ is an insufficient term, that ‘2nd and 3rd (inanimate) person forms also follow *mini* and *ni* “and/with”’ (§7). The statement is confusing and incomplete and not supported with examples; but, from its context, we may assume he was referring to forms like *m mini ba* and *m mini ya*. Nonetheless, he does not pursue this point and his term ‘after-verb’ pronouns does not account for such data.

3.2 Qualified Pronouns

Another consideration that causes problems for the preverbal/postverbal analysis of nonemphatic pronouns is the patterning of pronoun forms when combined with some qualifier, such as *zaa* (‘all’) or *kɔnko* (‘only/alone’). When the pronoun is followed by one of these qualifiers, it is the so-called ‘preverbal’ form that is used, regardless of grammatical role or position in relation to the verb, as illustrated in (4).

- (4) a. *ti puhiri be zaa ni be kookali*
 1P-A greet-IMPF 3P-A all CONJ 3P-A good.effort
 ‘We thank all of them for their good effort.’
- b. *di tu ni yi laxim yi kɔnko*
 3S/INAM-A be.proper COMP 2P-A unite 2P-A alone
 ‘You should get together yourselves [and not call on me].’
- c. *o kpuxi be niriba ayi maa bavili ba*
 3S-A take 3P-A people two DET straddle 3P-B
 ‘He gave the two of them a ride.’ [*niriba* serves as a dummy noun]
- d. *boomiya di gba biela*
 reduce-IMPER 3S/INAM also small
 ‘Reduce it also a little.’

Again, the analysis based on pre/postverbal position would have incorrectly predicted forms such as *ti puhiri *ba zaa*, *yi laxim *ya kɔnko*, *o kpuxi *ba niriba ayi maa bavili ba*, and *boomiya li* gba biela*.

Table 4 summarizes the findings of the variable set of nonemphatic pronouns when observed to combine with some sort of qualifier. The ‘qualifiers’ examined systematically for this analysis are numerals (which typically must also combine with a dummy noun (*zaɣ*’, *dib*’, or *niriba*), *kɔnko* and its truncated form *ko*’ (‘only’), *kam* (‘each’), *zaa(ha)* (‘all’), *gba* (‘also’), and indefinite articles (*sheli*, *sheŋa*, *so*, *sheba*—inam. sing, inam.

plur., anim. sing., anim. plur., respectively). Also included are a few unique instance of a pronoun plus noun (*yi pukpariba* ('you farmers') and *bε bihi* ('they children')); unfortunately, the corpus as coded does not facilitate a systematic verification of the existence or nonexistence of counter-evidence of similar constructions involving postverbal (B) forms of nonemphatic pronouns.

	PRO-A	PRO-B
PRO + qualifier + Verb	All cases (except as noted in next column) as predicted by any theory/account.	1 (1 <i>ti mini ba zaaha</i>), because of conjoined NP <i>zaaha</i> must have scope over the entire NP 'preverbal' classification would have strictly predicted <i>ti mini *bε zaaha</i>)
Verb + PRO + qualifier	28 (1 <i>n kɔnko</i> , 2 <i>yi zaa</i> , 1 <i>yi kɔnko</i> , 4 <i>bε zaa(ha)</i> , 1 <i>bε niriba ayi</i> , 2 <i>bε bihi ayi</i> , 1 <i>bε kam zaaha</i> , 8 <i>di zaa(ha)</i> , 4 <i>di gba</i> , 4 <i>di ko</i>) Plus 24 in written (2 <i>yi zaa(ha)</i> , 1 <i>yi gba</i> , 1 <i>yi pukpariba</i> , 2 <i>bε zaa</i> , 1 <i>bε niriba ayi</i> , 1 <i>bε gba?</i> (VD), 1 <i>bε yino kam</i> , 9 <i>di zaa</i> , 1 <i>di gba</i> , 1 <i>di sheli</i> , 1 <i>di sheŋa</i> , 2 <i>di ko</i> , 1 <i>di dib' ayi</i>)	6 (2 <i>ma gba</i> , 1 <i>ba gba</i> , 1 <i>ba zaaha</i> , 1 <i>mini ya zaaha</i> , [see notes on constituency in Section 5.1], 1 <i>li zaa</i>) Plus 8 in written (1 <i>mini ba zaa</i> , 5 <i>li zaa</i> , 2 <i>li dib'</i>)

Table 4. Nonemphatic Pronouns with Qualifier/Determiner (targeted cases).

Excluded from the list of 'qualifiers' is one of the most common, *pam* ('many/much'), whose dual grammatical roles (having either scope over an NP as a quantifier or over a clause or VP in a broader adverbial manner) cannot be disambiguated when following a nonemphatic 'postverbal' (B) pronoun, as well as the lexical item *mi* (roughly meaning 'too/also'), the problematics of which are too complex and are mentioned later (in Section 5.2). There were no instances of nonemphatic 'preverbal' (A) pronouns with *pam* in postverbal position, which might have further supported this analysis. Equally ambiguous is the particle *gba* which can have scope over either a specific NP or a broader action. Since the number of instances of pronoun + *gba* is relatively small, the results of this tally are included. *Zaa* has a similar dual role (as quantifier 'all' and adverb 'completely'; however, the intended role is usually more transparent than that of *pam* or *gba*. Nonethe-

less, some instances of the counter-evidence in Table 5 may in fact be instances of *zaa* and *gba* used ambiguously in an adverbial manner.

The results in Table 4 overwhelmingly discount any theory that describes Dagbani nonemphatic pronouns strictly in terms of position in relation to the verb. The only pieces of data in the spoken corpus that fit the prediction of the pre/postverbal account are a handful of instances involving *gba* and *zaa*. As noted above, the fact that these cases deviate from the general pattern (in which pronouns from series-A combine with qualifiers regardless of position or grammatical role) may be due to *gba* and *zaa* fulfilling an adverbial role rather than one of a nominal qualifier.

4 Towards a New Description

The evidence presented in this paper indicates that the idiosyncratic descriptions of Dagbani nonemphatic pronouns with terms like ‘postverbal’ or even ‘before nouns and verbs’ are hardly more successful in capturing the data than the ‘subject/object’ distinction. As long as we are using terminology which fails to capture all of the data, should we revert to the equally erroneous but more familiar ‘subject/object’ pronoun terminology? The descriptions employing terms like ‘preverbal’ and related expressions are improvements and do point in the right direction. The problem seems to stem from dwelling on descriptions that speak in terms of grammatical structures and lexical categories; whereas the phenomenon is probably best understood in phonological terms, or in the interface between syntax and phonology. In my initial analysis of this phenomenon, I proposed a fairly radical solution involving underlying monosegmental forms. Subsequently, I have pursued a variety of solutions, none of which are completely unproblematic when it comes to formalizing the rules. In what follows, I present various considerations in solving the puzzle of determining a concise, accurate, and comprehensive description of variant Dagbani personal pronouns. I leave this as a work in progress. For the sake of consistency I will retain the terms ‘preverbal’ and ‘postverbal’ throughout this article. However, while the complete details of a new solution remain to be worked out, I ultimately propose that the allomorphs of the nonemphatic pronouns be described as ‘proclitic’ versus ‘enclitic’ (or nonproclitic) forms, and that the category of nonemphatic pronouns can be listed as one set of pronouns with its allomorphs derived by a predictable set of rules or constraints. Furthermore, I would propose that the alternate terms, ‘weak’ and ‘strong’ rather than ‘emphatic’ and ‘nonemphatic’, be employed. While perhaps sounding less technical, these terms would better capture the fact that variation within the set of ‘nonemphatic’ pronouns is primarily motivated by low phono-

logical prominence, as opposed to a functional category like emphasis. For example, as noted elsewhere in this paper, one form of relative clause construction is drawn from the so-called emphatic pronouns, even though the head of the relative construction is not necessarily being emphasized in a functional sense.

4.1 Syntactic Solution: Operator/Operand; Predication Phrase

Although I claim that the best solution will take into account the apparent motivation of phonological reduction, it may in fact be possible to account for this phenomenon concisely in syntactic or structural terms (see also Section 5 on the syntactic-phonological interface related to pronoun variation). For example, the shared variation of ‘preverbal’ and possessive pronouns can be explained in terms of Predication Phrase (see Bower 1993), in which case we could refer to ‘subject’ (in the abstract sense) versus predicate forms. Similarly, Naden (1986) refers to associative/subject operator versus object operand when describing variant forms of personal pronouns in the Western Oti/Volta language family of which Dagbani is a member. But can a syntactic characterization like this incorporate the conjoined NP and qualified pronoun data?

4.1.1 Syntactic Solution Account of Pronouns in Conjoined NPs

On one hand, there are universal reasons to structure NP conjunction in a way that puts the first N/NP/Nominal Group of a conjoined NP in a Spec (hence, ‘Subject’ position) in X’ or PrP (Predication Phrase) structure (see Yoon & Lee 2005). Figure 1 fits an exemplary conjoined NP from Dagbani to ‘&P’ structure drawn from Yoon and Lee, which they pattern after Predicate Phrase structure.

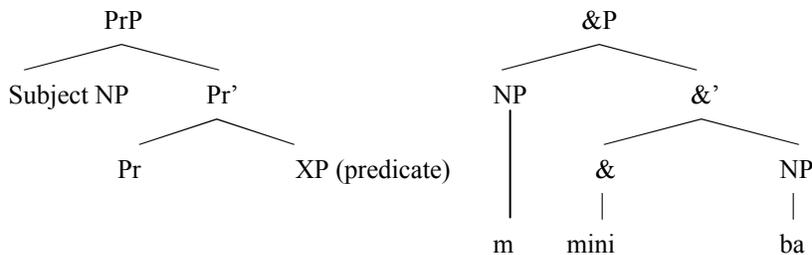


Figure 1. ‘&P’ Structure based on Predicate Phrase

Furthermore, on a more language specific level, Lord (1993: Ch. 4) analyzes NP conjunction in several neighboring Kwa languages (a sister language group to the Gur group), as well as a couple more remotely related Benue-Congo languages, as historically deriving from verb serialization

involving verbs meaning ‘be with’, ‘(bring) along with’, or possibly ‘take’ (e.g., Twi *ne*, Awutu *nɛ* Engenni *naa*). In such an analysis, the abstract &P could be replaced with or understood in terms of a more typical IP predicate structure.

Unlike the Kwa cases, there is no obvious connection between the conjunction (*mi*)*ni* and any known historical or defunct verb. However, there are parallel patterns of verb serialization in every other way; and Lord considers the possibility that syntactically and phonologically similar morphemes in Gur languages may derive from a similar process, making reference to a scholarly contention that ‘there are not good reasons to make Gur languages a separate classification from Benue-Kwa’ (ibid.: 51).

4.1.2 Syntactic Solution Account of Qualified Pronouns

Many of the ‘qualifiers’ I have identified could be analyzed as nouns. Certainly, the appositional *bɛ niriba ayi* (‘the two of them’; lit., ‘they two people’) could be analyzed as a case of two nominal phrases in an associative relationship structurally similar to a possessive phrase. Also, as with grammatical categories derived from verbs, Dagbani (as well as related languages) has many grammatical categories derived from nouns or possibly still functioning as nouns. The word *kɔŋko*, for example, is arguably a nominal meaning ‘oneliness’ (Naden, p.c.). The quantifiers like *zaa* on the other hand have too many parallels with the determiner/information status particles *maa* and *la*; and any nominal status would more likely have to be the result of ellipsis/reduction (as appears to occur in English, for example, in structures like ‘many have tried’ or ‘all are accounted for’). Attributing nominal (or lexical head status of any kind?) to *gba* is more problematic. The nature of this particle in general is unclear, but I doubt anyone, whether on intuitive or theoretical grounds, would consider this acting in a nominal capacity.

4.2 Phonological Solution: Proclitic Versus Enclitic Pronouns

As indicated earlier, based on a variety of considerations, I believe the best description will recognize the variant forms of Dagbani nonemphatic pronouns as being phonologically motivated, even if a viable description could be made in structural terms.

4.2.1 Comparison with Emphatic Pronouns

First, recall the less interesting variation of the emphatic or strong set of personal pronouns from Table 1. This involved a straightforward case of phonological reduction in the form of truncation or elision. Recognizing the

variation within the nonemphatic set of pronouns as a case of phonological reduction would establish a logical parallel between the variations for both types of personal pronouns.

4.2.2 Crosslinguistic Comparison

Crosslinguistically, the majority of sister languages in the Western Oti/Volta group only exhibit variant forms with the 1S category, and the weak/nonemphatic pronouns seem to be derived historically from noun class suffixes (hence, a single form) (Naden 1986). So, it makes more sense to hypothesize that Dagbani's variant forms are a phonological innovation, as opposed to saying all the other related languages have lost a structural contrast—the source of which has no bearing to the well-founded hypothesis of derivation from noun class suffixes. Furthermore, Dagbani is among the least preservative of proto Gur.

4.2.3 Clitics in Dagbani?

In addition to relatively unproblematic word classes such as noun and verb (but see Olawsky (2004) on problems of word classification), Dagbani is characterized by the existence of a variety of 'particles' (time depth markers, tense markers, emphatic particles) and 'suffixes' (emphatic, focus), as they are typically called in the absence of a more explicit grammatical description (Olawsky 1999, Wilson 1973). In a recent article on the elusive concept of 'word' in the world's languages, Olawsky (2002) makes an in-depth consideration of whether the term 'clitic' can be used to describe various lexical and morphological items in Dagbani such as these particles, based on generally accepted factors contributing to classification as clitics: lack of stress and phonological salience, attaching to a phonological host, variability of word order or allowing intervening materials between a potential phonological host, and inability to occur in isolation. While many of the particles end up failing one or another of the parametric tests for prototypical clitic status, such as the frequently stress-bearing time depth markers, Olawsky is fairly decisive in concluding that 'nonemphatic pronouns ... can be characterized as clitics' (218).

In his analysis, however, he focuses on 'preverbal' nonemphatic pronouns and only briefly refers to the identical possessive set (i.e. 'prenominal') and the 'postverbal' set. So, it is unclear whether he intends to classify 'postverbal' nonemphatic pronouns as 'enclitics'. Similar to the particles referred to as 'postverbal emphasiser' in Olawsky's article, the 'postverbal' pronouns are somewhat deficient as clitics in terms of their phonological salience by containing [a] as opposed to muted [ɛ] held by the 'preverbal'

counterparts. More importantly, I would argue that nonemphatic pronouns in object position do bear some stress.

Wilson (1973) also seems to have understood the phenomena of Dagbani pronouns more clearly than is suggested by the ‘preverbal’/‘postverbal’ dichotomy, the development of which presumably can be traced to his manuscript. Wilson was essentially writing for language learners, not linguistic theorists.⁴ As one of the first contemporary documented sources of Dagbani grammar, however, this manuscript set a precedent for linguistic descriptions of Dagbani pronouns based on position in relation to the verb. In light of the evidence presented in this paper and in combination with related linguistic phenomena in Dagbani—such as clitic-like behavior involving other lexical categories and related compounding processes (see Olawsky 2004)—I propose that the variation in nonemphatic pronouns is phonologically motivated by position of a clitic to its host. This would corroborate the observation of the truncated forms of the emphatic pronouns occurring when they appear before another word—i.e., also motivated by related combining or compounding phenomena. As seen in the data presented in previous sections, the phonological host for nonemphatic pronouns can be:

verbs in the case of subject and object functions,

nouns in the case of possessive function,

and more rarely when a noun is used as an adpositional qualifier as in *yi pukpariba* (‘you farmers’) or *bε bihi ayi* (‘these two children’, lit. ‘they/them two children’)

the **nominal conjunction** *mini* or *ni*,

and certain **qualifiers** (*kɔnko*, *ko*, *zaa(ha)*, *gba*, and various adpositional nominals including nouns and dummy nouns combined with numerals and, presumably, adjectives)

4.3 Underlying Forms

Whether we describe the variation in structural terms as in 4.1 or adopt a phonological description as in 4.2, the major problem in this puzzle involves the selection of the underlying set of pronouns and the specific rules or constraints that explain all of the alternations.

⁴ Other works by Wilson (1971a, 1971b) are more targeted towards an academic, linguistic audience; however, I mainly refer to Wilson (1973) in this paper as it is much more widely distributed and is arguably the dominant source of the ‘preverbal’ descriptive terminology.

4.3.1 Monosegmental UR Forms

The most radical solution might be to propose an underlying set of monosegmental preverbal forms (/m/, /a/, /o/, /d/, /t/, /y/). However abstract, such a solution is appealing due to corroboration with evidence and theoretical claims elsewhere related to the notion of monosegmental consonant morphemes in Dagbani and related languages. Rather than phonological reduction, this would imply a case where the underlying forms are of a minimal phonological form and rules relating to wellformedness call for the insertion of a nucleus where necessary, realized minimally as [ɿ] in proclitic position, but requiring a more prominent nucleus in enclitic position.

Olawsky (2002) hints at the possibility that the underlying form of certain pronouns may be a nonsyllabic consonant, such as /d/ for 3rd person singular inanimate. This notion is also suggested by Naden (ms.), who claims that such an analysis would support a reanalysis of the inventory of vowels in Dagbani (currently troubled by the issue of schwa or muted-[i]) to be more closely comparable with that of its sister languages. Furthermore, Naden also notes that ‘in Moore and Kusaal (and Bimoba of the Gurma group) some pronouns are analysed and written as a single consonant’ (ibid.: 4).

Under such an analysis, given their feature of [+syllabic], there would be no constraints impeding the direct attachment of /m/, /a/, and /o/ to a phonological host. So, these pronouns would attach as is with no epenthetic insertion. Only /m/ would adjust homorganically to the appropriate place of articulation. The nonsyllabic pronouns /d/, /t/, and /b/, however, would have to take an epenthetic vowel [ɿ] in order to attach to their phonological host, thus giving us the ‘proclitic’ forms [dɿ], [tɿ], and [bɿ], which we find written as *di*, *ti*, and *bɛ*. The 2nd person plural pronoun is a bit problematic.

In terms of the pronoun realization when a nonemphatic pronoun follows a host lexical item, the vocalic pronouns /a/ and /o/ would already be sufficiently prominent. The remaining nonsonorant forms, however, would take a prominent vowel that is most closely associated with its place of articulation. The bilabials /m/ and /b/ and the glide /y/, which already has a partially open oral passage, might be said to be well suited to take a highly ranked [a], thus giving [ma], [ya], and [ba]. The alveo-palatal /t/ and /d/, however, would end up taking [i] due to a rule or constraint that gives priority to proximity of place of articulation.

Such an analysis suffers from many of the considerations discussed in the subsequent subsections (treating the marked /m/ as underlying or default as opposed to /n/; predicting [yi] versus [yɿ]; and especially predicting [a] versus [i] as the nucleus for enclitic forms).

4.3.2 Preverbal Forms as UR

The simplicity principal as well as apparent metalinguistic view of native speakers (such as Yahaya (1995) implying preverbal as default) suggests treating preverbal forms as underlying, as these occur in a greater variety of environments. Compared to the monosegmental solution, this avoids the problem of accounting for the full vowel nucleus in [yi]. However, other problems mentioned for the monosegmental solution remain—namely, justifying marked allomorph /m/ as the UR form for the 1S form and explaining the realization of [i] versus [a] nuclei for the postverbal forms.

4.3.3 Homorganic Nasal Consideration

Any solution deriving postverbal forms from preverbal forms or from monosegmental forms suffers from having a more marked [m] as opposed to [n] serving as the default form of the homorganic nasal /ŋ/ 1S pronoun. Where the underlying form is a homorganic nasal /ŋ/, markedness theory would lead us to expect the unmarked form [n] to show up in default positions where place of articulation assimilation is not relevant. Indeed, we find phrases like *n yaba* ('my grandfather') realized as [n yaba] or even *n wumya* ('I have understood') as [n wumya]. However, if the postverbal form [ma] is hypothesized to derive from the preverbal form, we would also predict *[na] as opposed to [ma].

4.3.4 Postverbal Forms as UR

In light of the preceding considerations, there are synchronic language-internal reasons for positing the postverbal forms as underlying. Additionally, historical and cross-linguistic considerations tend to support a solution in which the postverbal forms are taken as underlying.

As mentioned under Section 4.2.2, personal and nonpersonal (in the case of inanimate nouns) pronouns in Dagbani and other languages from the Western Oti-Volta group appear to be derived from noun class suffixes (Naden 1986). This is especially apparent with languages other than Dagbani which have a full system of concordance between pronouns, as well as other anaphora, and noun class suffixes for up to 6 different singular/plural noun class pairs of inanimate nouns. These noun class suffixes, as word-final elements, naturally appear in a form similar to the postverbal forms of the personal and nonpersonal pronouns.

However, while this is very likely indicative of the origins of these pronouns, many linguists will be averse to letting diachronic information unduly influence the interpretation of synchronic phenomena. Furthermore, there remains the chicken or the egg dilemma by which the noun class suf-

fixes may have derived from pronouns. Naden also admits that ‘it is at least possible, however, ...to consider the more elaborate systems [of noun class/pronoun concordance pairs] as an analogic innovation, tidying up the system of noun-suffixes and partially-similar pronouns’ (ibid.: 262).

A solution taking the postverbal forms as underlying overcomes the problem of explaining [a]~[i] alternations in the nuclei of the postverbal forms, as well as avoiding the need to derive the [m] in *ma* from an underlying homorganic nasal. The main problem for this solution is the explanation of [yi] versus *[yɿ] for the 2P form.

4.3.5 The [yi]~[ya] Dilemma

In all solutions alternation between the 2P forms [yi] and [ya] poses certain problems. For the monosegmental and postverbal form solutions, which need to derive [yi] from some other underlying form, it is possible to claim that the relatively sonorant /y/ requires a fuller vowel nucleus to respect sonority hierarchy. Unfortunately, this is contradicted by the existence of the phonetic sequence [yɿ] in other lexical items such as the verb [yɿri] (‘fly’). Therefore, these solutions are complicated by the need to set the sonority hierarchy as a parameter that applies in certain contexts and not universally in the language.

In terms of predicting [ya], the monosegmental and preverbal form solutions require rules or constraints that support the notion that /y/ or /yi/ requires [a] in the nucleus in the postverbal position, despite the fact that the glide [y] is articulatorily well adapted to [i] and despite the fact that [i] is an acceptable nucleus for other postverbal forms such as [ti] and [li].

4.3.6 The [dɿ]~[li] alternation

The alternation between the 3rd person inanimate pronoun forms, [dɿ] and [li], is especially problematic for the solutions taking monosegmental and preverbal form as underlying. On the one hand, the lenition of /d/ to [l] would be somewhat contrary to the notion that the derived postverbal forms need to be more phonologically prominent. The inclination is simply to attribute this to an entirely different phenomenon—a general process of intervocalic or postvocalic allophony. In normal speech, even the proclitic *di* is weakened to a form with a liquid onset (usually heard as [r]) when falling in a postvocalic environment. Unfortunately for this solution, however, the proform [li] also occurs after a nasal, which is not typically a conditioning environment for allophonic lenition of [d] to [l]. In fact, in the case of the verbal suffix /-da/ or /-di/ which has [-ra] or [-ri] as a postvocalic allomorph, the /di/ is realized as [di]. Any description of nonemphatic pronoun

variation in Dagbani has to account for minimal pairs such as *wum li* ('hear it'), which is phonetically realized as [wum li], and *wumdi* ('hear-IMPERF'), which is pronounced as [wumdi].

In the case of the postverbal UR solution, the sonority hierarchy, which is only partially successful in explaining the realization of the 2P preverbal form as [yi] as opposed to [yɛ], could be called upon to help explain the alternation from /li/ to [dɛ]. In the case of [yi], the sonority hierarchy rule or constraint affects the realization of the nucleus, possibly because there is no natural nonsonorant counterpart for [y]. In the case of *[ɛ], the sonority hierarchy would be satisfied by the onset changing from the sonorant liquid /l/ to [d], which is presumably less sonorant than the minimally prosodic [ɛ] nucleus. In this regard, the postverbal UR solution mainly suffers in crosslinguistic terms, as the sister languages have [di] as the invariable pro-form and [-di] as the noun class suffix. In crosslinguistic terms, then, it might be preferable to modify the postverbal UR solution to have /di/ as an abstract UR for third person inanimate category. However, this then poses the problem of mentioned above of explaining the alternation from /d/ to [l] in allomorphic terms that apply in the enclitic environment (as in *wum li*) but not in the context of verbal inflection (as in *wumdi*).

4.3.7 The Inanimate Plural Form [ɲa]

Described in most grammatical descriptions as 'archaic' or 'becoming obsolete,' it is not particularly necessary to account for *ɲa* in a synchronic analysis. Indeed, the only occurrence in spoken texts in this corpus is in a fairly old recording of a traditional *salima* (fable) told by a seasoned fiddler and storyteller; but it *is* employed in new written texts as well as archaic ones. Based on the contexts in which it appears (translation of Koran, Ghana Constitution, and UNHCR Declaration; academic writings; technical documents), I would characterize it as 'formal' or 'literary,' but not 'archaic' or 'obsolete.' However, the fact that this form is said to be becoming obsolete, particularly in spoken discourse, may support the notion that the variant forms in Dagbani are an innovation. The emerging obsolescence of this form might be due to the impossibility of phonological reduction in preverbal position without ambiguous homophony with the 1S pronoun.

If relevant, however, all solutions would have the dilemma of explaining the archaic 3rd person plural inanimate form [ɲa]. In terms of crosslinguistic evidence or the correlation with noun class suffixes, *di* is paired with a separate plural counterpart (typically *a*, but sometimes *ɲa*), and there are additional remnants of a [li]~[ɲa] correspondence in Dagbani, as with the

inanimate indefinite pronouns *sheli* and *sheŋa*. If incorporated as a member of the nonemphatic pronouns, this would especially pose a problem for the monosegmental UR solution and the postverbal UR solution, both of which might appeal to a lexical exception on the grounds that the form [ŋa] is invariable in order to differentiate it from the velar allomorph of 1st person singular /ŋ/.

A more likely explanation, however, is that *ŋa* as documented and as employed in formal texts has been misclassified in a forced attempt to balance the set of singular-plural pronoun pairs, and simply does not form part of the nonemphatic pronouns in terms of the synchronic development of the spoken language. In current dictionaries and wordlists, it is given glosses like ‘these’, ‘that’, ‘those’, ‘(several things)’, and not as ‘it’. Observing, like most everyone else, that the word is ‘becoming obsolete,’ Mahama (2003) provides the demonstrative pronouns *dinaa* [sic] and *dinnima* as ‘the current words,’ not a nonemphatic pronoun like *di*. Wilson (1973) suggests that *ŋa* is an archaic emphatic pronoun, and he has this to say about the plural form of 3rd person inanimate pronouns:

there is no difference between singular and plural with *dɛ*, *li*. (In the NT [New Testament] you will find an old strong inanimate plural pronoun *ŋa*, but it is very little used; it need only be recognized, not imitated; though some people may try to tell you *dɛ*, *li* cannot be plural, this is somewhat over-pure for everyday usage.) (§7)

Classifying /li/, /di/, or /d/ as fulfilling both singular and plural functions would be consistent with the fact that, with the exception of the idiosyncratic constructions of the form *man' gbazaa*, *a gbazaa*, etc. (roughly, ‘all of me’ or ‘me completely’, and ‘you completely’), in instances where the qualifier *zaa* (‘all’) clearly has scope over a solitary nominal item, the only attested forms are plural pronouns (*ti*, *yi*, *bɛ*, *yinim'*, *ban'*, *dinnima*), plural nouns or nouns implying plurality like *gaana* (‘Ghana’), and other nominals implying plurality like *bɛ yino kam zaa* (‘each and every one of them’), and many instances of *di* in a variety of contexts. Also, in subject-head relative constructions, in which a truncated strong (‘emphatic’) pronoun follows with an optional indefinite article, *din* is frequently combined with the plural inanimate indefinite article *sheŋa*, in lieu of *ŋan*.

5 A Note on the Interface Between Syntax and Phonology

5.1 Cliticization and Constituency Boundaries

Having analyzed the variation in nonemphatic pronouns in Dagbani as a process of cliticization, I would like to address what I interpret as an erroneous claim about Dagbani ‘clitics.’ In speaking of the degree of ‘selectivity’ that clitics in a given language may assert in terms of the word class of their host, Aikhenvald (2002) cites Dagbani as an example of a language with clitics having ‘low selectivity.’ In a misinterpretation of Olawsky’s chapter in the same volume, Aikhenvald states that ‘all clitics are attached to a host (defined phonologically) with which they may not have a grammatical connection’ (ibid.: 45). First of all, the exact status of Dagbani particles is not straightforward. Olawsky (2002) has taken a major, insightful first step in considering which of lexical items in Dagbani may qualify as ‘clitics,’ but linguists working with Dagbani do not decisively refer to any given ‘particle’ or lexical item as a ‘clitic.’ Olawsky himself cautions, for example, that the classification of preverbal particles as ‘clitics’ is ‘more controversial.’ And his conclusion on the subject rather refers to a spectrum of sorts between word and affix. He concludes, ‘Dagbani has a variety of clitics, which may have different properties. Some are obviously closer to affixes, whereas others share more features of words’ (ibid.: 223).

As the data in the paper show, the hosts of pronoun clitics are not simply ‘defined phonologically.’ Phrasal constituency allows us to accurately predict the attested forms where there might otherwise be some ambiguity or indecision based solely on a potential adjacent host. It also helps to explain some surprising occurrences, including some of the unpredicted forms presented in the tables in preceding sections. For example, the ‘preverbal’ description would definitely not explain the sentence in (5).

- (5) [ti mini ba] zaaha wɔlisi li -mi je
 1P-A CONJ 3P-B all struggled.with 3S-B EMPH be.weary
 ‘We and they all struggled with it in vain/tirelessly.’

However, a clitic-based description relying solely on phonological considerations would not necessarily predict this form either, as the 3rd person plural pronoun could just as well cliticize with the qualifier *zaaha*. In fact, there might even be reasons to predict this as the more natural result based on phonological patterns alone, if procliticization seems to be more common in Dagbani. However, the attested ‘enclitic’ form can be understood based on phrase structure and constituency levels. The qualifier *zaaha* has scope over the entire conjoined NP *ti mini ba*. Thus, in this case, the pronoun /ba/ (or whatever the underlying form) combines with *mini* within the NP, and the qualifier *zaaha*, combining with the conjoined NP at a higher phrase level, has no bearing on the realization of /b/:

[[[ti]_{NP} mini [ba]_{NP}]_{NP} zaaha]_{DP}.

Hypothetically, we might expect a minimal pair in which the qualifier only has scope over the one pronoun at a lower level prior to the NP conjunction: [[ti]_{NP} mini [[bε]_{NP} zaaha]_{DP}]_{NP}. However, no such data was found in my corpus, and it would be hard to come across or even contrive the right context to elicit this form.

A similar instance, where the conjoined noun phrase serves a possessive function rather than verbal subject, is presented in (6).

- (6) ti dihiri [ti mini ba] zaa puya
 1P feed-IMP 1P CONJ 3P-B all stomachs
 ‘We feed all of our and their stomachs’

As with (5), in (6) we see *ba* as opposed to *bε*, because the qualifier *zaa* has scope over the entire NP:

[[[[ti]_{NP} mini [ba]_{NP}]_{NP} zaa]_{DP} puya]_{NP}.

Again, though theoretically desirable, there are no attested minimal pairs or near minimal pairs showing a case where *zaa* has scope only over the second pronoun. However, the next set of data in (7) involving a comparable construction does include such minimal pairs.

The sentence in (7a) shows a similar construction minus the qualifier. Here, we see the constituency level consideration applying with nonemphatic pronouns in a conjoined NP which is in a possessive construction with an adjacent noun (actually, there are several embedded possessive constructions in this example).

- (7) a. di nye -la [a mini ba] tuuli taba nyabu
 3S/INAM-A COP FOC 2S CONJ 3P-B first REFL seeing
 ‘It is your and their first time meeting each other’
 b. [bε_i] mini [bε_i bieli maa] zaa daa yi na
 3P-A CONJ 3P-A elder.sibling DET all TD2 exit hither
 ‘They and their elder brother all went out’

Since both pronouns in the conjoined pronoun NP are intended as possessors for the possessed noun, the 3rd person plural pronoun takes the conjunction *mini* as its host and not the nominal item that follows:

[[[a]_{NP} mini [ba]_{NP}]_{NP} [tuuli[taba[nyabu]_N]_{NP}]_{NP}.

Near minimal pairs of this sort are easily attestable, as in (7b). In this sentence, the first 3P pronoun in the conjoined NP functions as a subject pronoun ('they') while the second pronoun functions as a possessive pronoun ('their'). Since the only possessor of the noun *biɛli* ('elder brother') is the second *bɛ*, this pronoun takes the noun *biɛli* as a host at the lowest constituency level and this possessive NP construction conjoins with the first pronoun at a higher level:

[[bɛ]_{NP} mini [[bɛ[biɛli]_N]_{NP} maa]_{NP} zaa]_{DP}.

If cliticization were strictly phonological, we might have expected the pronoun to take *mini* as its host, giving *bɛ mini *ba biɛli maa zaa daa yi na*.

5.2 Interaction with Other Clitics/Particles

One consideration which may have influenced Aikhenvald's categorization of Dagbani clitics as having low selectivity is the fact that clitic pronouns can actually appear before lexical items other than the list of potential hosts summarized in Section 4.2.3. This includes various preverbal particles, included among the lexical items Olawsky considers as potential candidates for classification as clitics. For example, in the complex nominal in (8), the subordinating particle *ni* (which is used in combination with indefinite articles, such as animate/plural *sheŋa* in this instance, in the formation of relative constructions such as this one) comes between the nonemphatic pronoun *di* and the verb *di* ('eat').

- (8) zaŋ kana ti ya zaa mini di **ni** di d'alifaani sheŋa mi
 take come 1P town all CONJ 3S/INAM-A SUB eat benefit INDEF/REL also
 'concerning our whole town and the benefits it has reaped' [written, letter]

In this instance, the subordinator *ni* (phonetically [nɛ]) is likely classifiable as a clitic with just as little phonological prominence as the nonemphatic pronoun, and the verb *di* is arguably the phonological host of both clitics. The fact that *ni* intervenes between the subject pronoun and its host is based on ordering rules, not some random selectivity on the part of the pronoun clitic. The whole sequence is pronounced as [dɔndi] or perhaps more like [dɔndi].

While the preverbal position is the most common environment where a clitic pronoun will be followed by an intervening lexical item, due to the extensive use of preverbal particles in Dagbani; we also see instances of intervening particles elsewhere. In (9), we see the particle *mi* ('also, too',

having a similar sense as *gba*, which we saw among the ‘qualifier’ hosts in Section 3.2) appearing between the weak pronoun *bε* and the qualifier *zaa*.

- (9) *be mi zaa daa niŋdi ɲɔ*
 3P-A also all TD2 do-IMPF DEM
 ‘they all also were doing this’

This particle *mi* was excluded from the list of potential ‘qualifiers’ in Section 3.2 due to a variety of complications, including, for example, its ambiguous relationship with the suffix *-mi*, or what Olawsky (1999) calls a ‘postverbal emphasiser.’ Another reason, however, was that an apparent limited distribution (strictly in proclitic positions when following a nonemphatic pronoun) suggesting it to be as much a dependent clitic as the nonemphatic pronouns themselves. In addition to examples like the sentence in (9), *mi* is found to intervene between nonemphatic pronoun and possessed noun, and between pronoun and verb. Similar to the example in (8), the vowels of both the pronoun and the particle are muted ([bɪmɪza:] or [bɪmɪzaa]). Hence, the host, be it qualifier, noun, or verb might be said to be the host of two clitics, with the intervention of *mi* between pronoun and host being explained by sequencing restrictions, not low selectivity by the clitic.

More difficult to explain is when a more prominent particle intervenes, such as the time depth marker *daa* (indicating that a time span of two or more days has passed or will pass between point of utterance and point of reference). Does such a particle qualify as the host (maybe an interdependent host) in this case, and does this imply low selectivity on the part of the preceding pronoun? Maybe so, but I believe the controversial, ambiguous status of the various particles in Dagbani need to be clarified before citing Dagbani as a typical example of this or that parameter of cliticization.

6 Conclusion

Perhaps at some point in the development of the Dagbani language, the description of proclitic pronouns as ‘occurring before nouns and verbs’ might well have been sufficient in capturing all data regarding the realization of nonemphatic pronouns. For example, perhaps the qualifiers, such as *kɔŋko* and *zaa* can be viewed to exhibit the qualities of nominals or to be derived from the same; and perhaps (though, even more radical of an idea) the nominal conjunction *mini* and *ni* derived historically from a serializing verb (or a verb **ni* preceded by the particle *mi*?) similar to the constructions for nominal conjunction in genetically related languages (Lord 1993). In synchronic terms, however, these have developed into unique word classes, and the patterning of nonemphatic pronouns in relation to these lexical items

needs to be incorporated into a comprehensive description of the Dagbani pronominal system and grammar in general. Despite the fact that the contrary evidence posed by these instances have been overlooked or received little explanation in grammatical descriptions, this is by no means a rare occurrence.

Given the problematics of defining just what it means to be a clitic in Dagbani or any language, the cautious, intermediary solution would be to simply describe the appearance of variable pronoun forms in bulky, conditional terms such as 'series A appears before nouns and verbs, except in such and such instances...'. However, I believe a theoretically sound, concise, and efficient description can be developed employing the cliticization-based analysis pursued in this paper. Furthermore, such a solution will likely corroborate other phenomena and theoretical issues in Dagbani grammar. Much more remains to be studied or clarified regarding various issues of cliticization in Dagbani and the degrees of classification of the various 'particles' in Dagbani. Hopefully, the additional phenomena presented and discussed in this paper will serve as a useful contribution to the resolution of such issues.

7 Abbreviations

- separates isolatable morphemes, suffixes
- /- lists inseparable grammatical information, fused morphology
- 1S/2S/3S – 1st/2nd/3rd person singular (*n, a, o, di, li, mani, nyini, ḡana*)
- 1P/2P/3P – 1st/2nd/3rd person plural (*ti, yi, ba, ḡa, tinima, yinima, bana, ḡana*)
- A – ‘pre-’ forms of nonemphatic pronouns (*n, di, yi, be*)
- ANIM – animate (used in glossing INDEFInite articles/pronouns: *so, sheba*; not glossed with personal pronouns)
- B – ‘post-’ forms of nonemphatic pronouns (*ma, li, ya, ba*)
- COND – conditional particle (*yi* - REALis, *naan* - IRREALis)
- CONJ – conjunction (*mini, ni* – nominal; *ka* - verbal)
- COP – equative, copular verb (*nyε*)
- DEM – demonstrative (*ḡo, dimbḡḡo, dimbala, dina, dinnima*)
- DET – determiner (*maa* – known/given information, *la* – shared/retrievable information)
- EMPH – emphatic (emphatic pronouns (*mani, nyini*, etc.); and emphatic function of INF)
- FOC – focus marker (*-la*, marker of new information)
- FRO – left fronting/focus construction (*ka*)
- FUT – future (*ni; ku* – NEGative)
- IMPER – imperative suffix (*-ma* (*-mi, -miya*))
- IMPF – imperfective (*-da* (*-ra, -di, -ri*))
- INAM – inanimate (*di, li, ḡa, dina, ḡana* - personal pronouns, ANIMATE forms are assumed default and not coded ANIM; also used in glossing INDEFInite articles/pronouns: *sh□li, sheḡa*)
- INDEF – indefinite article/pronoun (*sheli, sheḡa, so, sheba*)
- INF – infinitive prefix (used as EMPHatic and in verb sequencing)
- LOC – locative (*ni* – postpositional/nominal; *be* - verb)
- NEG – negative (*bi, ku*)
- NP – noun phrase
- PLUR – plural (when glossing INDEFInite articles/pronouns: *sheḡa, sheba*; nominal suffix *-nima*, glossed as separate item, because it can inflect complex NPs (comparable to English ‘s))
- PRO – pronoun
- TD0/TD1/TD2 – time depth markers (*di* – same day, *sa* – 1 day, *daa* – 2 or more days)
- RECIP – reciprocal (*taba*)
- RECIP – reflexive pronoun (*maḡa*)
- REL – relative (a function filled either by EMPHatic pronouns *ḡun, ban, din, ḡan*; or by INDEFInite pronouns in combination with SUBordinator *ni*)
- SING – singular (used when glossing INDEFInite pronouns: *sheli, so*)
- SUB – subordinator (*ni*)
- TOP – topicalizer (a function of *maa*, which is crudely glossed as DETerminer)
- V - verb
- VP – verb phrase
-
- STR – strong pronouns (proposed gloss in lieu of EMPHatic for personal pronouns, reserving EMPHatic for other emphatic functions, following the analysis proposed in this paper; weak pronouns are default)

8 References

- Aikhenvald, A. Y. 2002. Typological parameters for the study of clitics, with special reference to Tariana. *Word: A Cross-Linguistic Typology*, eds. R. M. W. Dixon & A. Y. Aikhenvald, 42-78. Cambridge: Cambridge University Press.
- Bawa, A. 1980. Collected Notes On Dagbani Grammar. Unpublished manuscript (revised edition). Ajumako, School of Ghana Languages.
- Bowers, J. 1993. The Syntax of Predication. *Linguistic Inquiry* 24:591–656.
- Lord, C. 1993. *Historical Change in Serial Verb Constructions*. Amsterdam: Benjamins.
- Mahamma, I. 2003. Dagbani – English Dictionary. Tamale, Ghana: Ibrahima Mahamma, School for Life.
- Mohammed, K. A. 2004. Handout on Ghanaian Languages and Culture for Students in Teacher Training Colleges. Unpublished manuscript. Tamale, Ghana: Dept. of Ghanaian Languages, BATCO.
- Naden, T. 1986. Western Oti/Volta pronoun systems. *Pronominal Systems*, ed. Ursula Wiesemann, 249-275. Tübingen: Günther Narr.
- Naden, T. ms. Sch-u-wa in Dagbani. Unpublished manuscript. Tamale, Ghana: GILLBT.
- Olawsky, K. J. 1999. *Aspects of Dagbani Grammar, with special emphasis on phonology and morphology*. Muenchen: LINCOM Europa.
- Olawsky, K. J. 2002. What is a Word in Dagbani? *Word: A Cross-Linguistic Typology*, eds. R. M. W. Dixon & A. Y. Aikhenvald, 205-226. Cambridge: Cambridge University Press.
- Olawsky, K. J. 2004. What is a Noun? What is an Adjective: Problems of Classification in Dagbani. *Journal of African Linguistics* 25:127–148.
- Wilson, W. A. A. 1971a. Class pronoun desuetude in the Mõõré-Dagbani subgroup of Gur. *Journal of West African Linguistics* VIII (2):79–83
- Wilson, W. A. A. 1971b. Verbal sequence and case markers in Dagbani. Unpublished PhD Dissertation. University of Texas at Austin: University Microfilms
- Wilson, W. A. A. 1973. *Dagbani: An Introductory Course*. Tamale, Ghana: GILLBT.
- Yahaya, R. M. 1995. *Dagbani Karɔŋ Piligu*. Accra: Bureau of Ghana Languages.
- Yoon, J. H. S. & W. Lee. 2005. Conjunction Reduction and its Consequences for Noun Phrase Morphosyntax in Korean. Cascadilla Proceedings Project. (www.linguistics.uiuc.edu/~jyoon/Papers/Yoon_Lee_May_4.pdf)

What Sorani Kurdish Absolute Prepositions Tell Us about Cliticization

POLLET SAMVELIAN

1. Introduction

Sorani Kurdish dialects have independent personal pronouns, clitics pronouns and verbal personal endings.

Independent pronouns			Clitic pronouns			verbal endings		
	Sg	Pl		Sg	Pl		Sg	Pl
1	<i>min</i>	<i>(h)êma</i>	1	<i>-(i)m</i>	<i>-mân</i>	1	<i>-(i)m</i>	<i>-în</i>
2	<i>to</i>	<i>êwa</i>	2	<i>-(i)t</i>	<i>-tân</i>	2	<i>-î</i>	<i>-(i)n</i>
3	<i>aw(a)</i>	<i>awân</i>	3	<i>-î/-y</i>	<i>-yân</i>	3	<i>-ê(t)/∅</i>	<i>-(i)n</i>

Table 1

Table 2

Table 3

Independent pronouns have exactly the same distribution as noun phrases and we will not consider them in this paper. Table (2) and (3) both contain bound forms which do not bear lexical stress and are always attached to a host, although only the former are regarded as ‘clitic pronouns’ because of their placement properties. Contrary to Table (3), affixes, which are always attached to a verbal host and hold a function with respect to the latter (as agreement marker or argument), the forms in Table (2) can occur in different positions in the sentence and do not exhibit any restriction with respect to their host. Furthermore, they do not occur necessarily in relation with a verbal head, but can also be located within a noun phrase or a prepositional phrase, as

a complement of the head noun or preposition. Thus, contrary to the forms in Table (3), which can be clearly identified as lexical affixes, the forms in Table (2) are reminiscent of a well-known class of special clitics, generally referred to as ‘second position’ (2P) clitics. Even though this is undoubtedly an oversimplification, let us assume that their cliticization domain is the VP and that they are attached to the right edge of the first constituent within this domain, as illustrated by the following examples:

- (1) Narmîn u Sirwan bâng=**mân** da-ka-n
 Narmîn and Sirwan voice=1.PL AM-do.PRES-3.PL
 ‘Narmin and Sirwan are calling us’
- (2) min ba Narmîn=**î** da-lê-m
 I to Narmin=3.SG AM-tell.PRES-1.SG
 ‘I am telling it to Narmin’
- (3) min ba Narmîn=**î** ba kurdî da-lê-m
 I to Narmin=3.SG in Kurdish AM-tell.PRES-1.SG
 ‘I am telling it to Narmin in Kurdish’

It is beyond the scope of this paper to provide a comprehensive description of clitic placement possibilities in Sorani. Whatever the exact placement rules for clitics are, attachment to the subject NP seems to be regularly excluded for pronominal clitics, as shown by the following example:

- (4) * Narmîn=**yân** da-kuj-ê
 Narmin=3.PL AM-kill.PRES-3.SG
 (putatively) ‘Narmin is killing them’

Apart from differences due to their placement properties, clitic pronouns and personal verbal endings hold the same functions with respect to the verb and are in complementary distribution in the following way:

- A.** With transitive and intransitive verbs in the present tenses and only intransitive verbs in the past tenses, personal endings realize subject agreement and are obligatory:

- (5) (min) kitêb-êk bo Narmîn **da-kir-im**
 (I) book-INDEF for Narmin AM-buy.PRES-1SG
 ‘I am buying a book for Narmin’
- (6) bo çî döne **na-hât-î** bo mâl-î
 for what yesterday NEG-come.PAST-2.SG to house-EZ

êma ?
 we
 ‘Why didn’t you come to our house yesterday ?’ (Blau, 1980, p.68)

Clitics, if present, are generally interpreted as the direct object of the verb:

- (7) (min) bo Narmîn=**î** da-kir-im
 (I) for Narmin=3.SG AM-buy.PRES-1.SG
 ‘I am buying it for Narmin’

B. With transitive verbs in the past tenses, a reversed distribution is observed. Clitics realize subject agreement, and are attached to the right edge of the VP’s first constituent:

- (8) (min) kitêb-êk=**im** bo Narmîn kirî
 (I) book-INDEF=1.SG for Narmin buy.PAST
 ‘I bought a book for Narmin’
- (9) bâzîrgân-akân asp-akân=**yân** da-kirî
 merchant-DEF.PL horse-DEF.PL=3.PL AM-buy.PAST
 ‘The merchants were buying the horses’ (Blau, 1980, p.71)

Personal verbal endings, if present, are interpreted as a direct object:

- (10) Narmin ba Sirwan=**î** dâ-**n**
 Narmin to Sirwan=3.SG give.PAST-3.PL
 ‘Narmin gave them to Sirwan’
- (11) ba dûrbîn=**mân** dît-**in**
 with binoculars=1.PL see-3.PL
 ‘We saw them with binoculars’

This implies that transitive verbs in the past tenses are not necessarily followed by personal endings, as illustrated by the examples (8) and (9) above.

This pattern is reminiscent of split ergativity, and has led Mackenzie (1961), among to analyze the transitive construction in the past tenses as an ‘agential construction’. Mackenzie claims that in this construction, the noun phrase referring to the Agent argument of the verb ‘is in no way equivalent to a Subject, in concord with the verbal form’ (p. 107). If the Agent is not the subject, it follows that the clitic cannot

realize subject-verb agreement, but is in fact the Agent argument of the verb. That is the reason why Mackenzie uses the term ‘agential suffix’ to designate the clitic in this case. Though it is not explicitly claimed by the author, such an analysis implies that the grammatical subject of a transitive verb in the past tenses is in fact its Patient, or its ‘Direct Affectee’ in Mackenzie’s terms. This would explain the use of verbal personal endings, which realize subject-verb agreement, in a similar way as with the present tenses.

The main advantage of Mackenzie’s view is that it provides a unified account for each set of personal bound morphemes: the forms in Table (3) are always regarded as inflectional verbal affixes and function as agreement-markers, while clitics realize an argument of the verb and constitute one of the options for argument realization, together with noun phrases (or prepositional phrases) and independent pronouns. In other words, they are bound pronouns, and not agreement markers. Appealing though it is, and supported by historical facts, this analysis faces serious problems.

First, in the past transitive construction, personal endings are generally in complementary distribution with overt noun phrases or overt independent pronouns. Although it is possible for the ‘Direct Affectee’ NP to be doubled by a personal verbal ending, as it is the case in the following example, this doubling is by no means obligatory:

- (12) *dû nâma=t ba kurdî nûsî(-n)*
 two letter=2.SG in Kurdish write.PAST(-2.PL)
 ‘You wrote two letters in Kurdish’

Viewing personal endings as agreement markers in this case would amount to posit an optional subject-verb agreement in the past transitive construction. Since subject-verb agreement is obligatory in Sorani dialects, this is problematic and Mackenzie himself acknowledges that it would be an over-simplification to state that the verb agrees with the Direct Affectee.

Second, a clitic referring to the Agent is always present in the past transitive construction, irrespective of the presence of a noun phrase or an independent pronoun referring also to the Agent. Thus, the following sentence, in which the clitic referring to the Agent has been removed, is ungrammatical:

- (13) * (min) *kitêb-êk bo Narmîn kirî*
 (I) book-INDEF for Narmin buy.PAST
 (putatively) ‘I bought a book for Narmin’

In other words, clitic doubling is systematic in this construction. This supports the view of the clitic as a subject-verb agreement marker, rather than as one of the available options for argument realization, on par with overt noun phrases and independent pronouns.

Finally, the Agent NP displays the prototypical properties of a grammatical subject. For instance, in the following coordinate structure, the missing argument of the past transitive verb *dîtin* 'see' is interpreted as the Agent and is coreferent with the subject of the intransitive verb *hâtin* 'come':

- (14) Narmin hât u Sirwan-î dît
 Narmin come.PAST and Sirwan-3.SG see.PAST
 'Narmin came and saw Sirwan'

From this body of evidence, it may be concluded that the Agent in the past transitive construction is also the grammatical subject of the sentence and that the clitic referring to the Agent realizes subject-verb agreement.

Under this assumption, if clitics are viewed as syntactic items, we are committed to admit that subject-verb agreement is realized by two distinct devices in Sorani Kurdish dialects, depending on the verbal construction and its tense: in the past transitive construction, subject-verb agreement is syntactically handled and realized by a clitic, while in all other cases, it is a morphological process, realized by an affix on the verb. From a conceptual point of view, this analysis is inferior to Mackenzie's unified account, where subject-verb agreement is always a morphological matter. One way of avoiding this drawback would be to adopt the alternative view of clitics as affixes, handled in morphology, and this is indeed the analysis it will be argued for in this paper.

Such an alternative view of clitics in Sorani may seem quite surprising at first sight. Indeed, the data examined up to now supports rather the opposite view, in that clitic placement has been determined with respect to a syntactic constituent (i.e. the VP). If their placement is always defined in syntactic terms, pronominal clitics may at best be regarded as phrasal affixes, but certainly not as lexical affixes. Even though a morphological treatment of phrasal affixes is conceivable and argued for in various studies, the phrasal affix analysis of clitics leaves us nevertheless with our initial problem, that is the realization of subject-verb agreement by two distinct categories of items: lexical affixes in the case of personal endings, and phrasal affixes in the case of clitics. The same situation stands for the pronominal realization of direct objects

Along the lines that follow, it will be first shown that pronominal

clitics are not unambiguously phrasal affixes and that in some contexts they behave very much like lexical affixes. Two contexts provide ample evidence in favor of a lexical affix analysis of clitics. The first concerns cliticization on the verbal host, the second one involves clitic complements of so-called absolute prepositions.

2. Clitics as affixes: the case of endoclitics

The first body of evidence comes from phenomena generally referred to as ‘mesoclitisis’ or ‘endoclitisis’ (Zwicky, 1977), involving the interruption of a word by a clitic. Since clitics attach to the right edge of the first constituent of the VP, in case the latter contains only a verb, the clitic is expected to occur after the verb. However this is not what happens, as shown by the following examples:

- (15) a. da=**m**-xwârd b. na=**m**-xwârd
 TM=1.SG-eat.PAST NEG=1.SG-eat.PAST
 ‘I was eating’ ‘I did not eat’
- c. na=**m**-da-xwârd
 NEG-1.SG-AM-eat.PAST
 ‘I was not eating’

In all these cases, the clitic is placed after the first prefix of the verbal inflected form. If the verbal stem is not preceded by a prefix, the clitic follows it, and thus intervenes between the latter and the personal ending:

- (16) a. dî=**t**-în b. nard=**man**-in
 see.PAST-2.SG-1.PL send.PAST-1.PL(SUBJ)-3.PL(OBJ)
 ‘You saw us’ ‘We sent them’

The generalization seems to be that the clitic is still in the second position in some sort, but "second position" is to be defined on morphological grounds, that is after the first morpheme (affix or stem) of the word. These facts pose a conceptual problem for a uniform analysis of clitics both as phrasal affixes or as syntactic items. Pronominal clitics occur within a word and are linearized with respect to other morphological items (stems or lexical affixes), including lexical affixes. Thus, if personal verbal endings are lexical affixes, pronominal clitics must also be regarded as lexical affixes in this case, and not as phrasal affixes, neither as syntactic items.

Relying on similar facts observed in Pashto, European Portuguese and Udi, Anderson (2005) assumes however that even in the case of

endoclitics, the phrasal affix analysis is the best option to retain. Anderson claims that endoclitics never interrupt genuine single lexical words and that all well-known cases of endoclitics involve in fact the placement of a clitic after the first lexical word of a morphologically complex word. One set of facts which seem to challenge Anderson's analysis is provided by the European Portuguese enclitics, which are comparable to some extent to Sorani data under discussion here. It is well-known that, in European Portuguese, when the verb is a form in the synthetic future or conditional, the clitic is located internal to the form between the stem and the ending, giving thus rise to forms such as the following:

- (17) mostr-ar=ilho=emos
 show-FUT=3.SG=1.PL
 'We will show it to him'

Given the fact that the inflected future and conditional, contrary to all other tense forms, result from the relatively recent fusion of a separate auxiliary verb (a form of *haver* 'to have') with the infinitive form of the lexical verb, Anderson assumes that these forms reflect their relatively recent reanalysis from originally complex forms and that the structure of, e.g. *dariamos* '(We) would give' is that in (18).

- (18)
-
- ```

graph TD
 V --> LexWord
 V --> Node1[]
 LexWord --> dar
 Node1 --> i
 Node1 --> amos

```

It is further assumed that the material in (18) is organized in a single Phonological Word, whose stress pattern is determined by that of its second element. When a clitic is added, if the verb is initial within the Intonational Phrase, the former, cannot be realized in leftmost position within the verbal domain. Given the fact that it must be positioned as close to the left edge of this domain as possible, the clitic will be positioned at the right edge of the lexical word, but preceding the conditional ending. In sum, under Anderson's account, pronominal clitics in European Portuguese are uniformly phrasal affixes positioned with respect to the VP. They do not interrupt lexical words and their placement follows from the general constraints on the second position. In other words, there are no endoclitics in European Portuguese.

If we refer to (15) and (16) above, such an analysis can be applied to Sorani clitics. Indeed, the latter never interrupt a simple lexical unit,

and always follow the first morpheme of a morphologically complex verb. Problems arise however when further data are taken into account, which go against the generalization on second position placement of clitics within a complex verbal form.

First, when the clitic is in the 3rd singular form, it must be placed after the personal ending:

- (19) nard-in=î  
 send.PAST-3.PL=3.SG  
 'He sent them'

Apart from this idiosyncratic placement, which occurs in all Sorani dialects, placement variations are observed from one dialect to another. Mackenzie (1961) notes for instance that in Piždar and Mukrî dialects, the string *-im-in* 'I...you/them' alternates freely with *-in-im*.

The past participle suffix *-u/-uw* constitutes another exception. Clitics, as well as personal endings, cannot interrupt the string composed of a verbal stem and the past participle suffix. The past participle is used in the formation of perfect tenses. The suffix *-u/-uw* is directly attached to the verbal stem and precedes both clitics and verbal personal endings:

- (20) a. nârd-uw=tân-in  
 send.PAST-PP=2.PL-3.PL  
 'you have sent them'  
 b. \* nard=tân-uw-in  
 send.PAST=PL.2-PP-3.PL

Once again, the clitic occurs in the third position. It could be objected that in this case there is a prosodic explanation to the third position placement. Indeed the lexical stress in (20) falls on the past participle suffix and not on the verbal stem. It could be assumed that the suffix and the verbal stem form a lexical word and that the clitic appears after the first lexical word, i.e. an accented unit, occurring thus in second position. The problem with this account is that it leaves unexplained the clitic placement with regard to the imperfective prefix *da-*. We saw already that when the verbal stem is preceded by an aspectual or modal prefix, the clitic intervenes between the prefix and the verbal stem. However the prefix *da-* is unaccented and the lexical stress in this case falls on the last syllable of the verbal stem: *da=m-xwârd* 'I was eating'. These facts support the claim that within a complex verbal form, the placement of clitics cannot be accounted for in terms of second position, whatever the definition of such a position be. Placement idiosyncrasies,

which are characteristic of lexical affixes, challenge not only the analysis of clitics as syntactic items but also a uniform analysis in terms of phrasal affixes.

The last body of evidence against both the syntactic analysis and a uniform phrasal affix analysis of endoclitics is provided by coordination. Let us consider the following examples:

- (21) a. bo Narmîn aw kitêb-a da-nus-im u  
to Narmin this book-DEF AM-write.PRES-1.SG and  
da-xwên-im  
AM-read.PRES-1.SG  
'I am writing and reading this book for Narmin'
- b. bo Narmin=**î** da-nus-im u  
to Narmin=3.SG AM-write.PRES-1.SG and  
da-xwên-im  
AM-read.PRES-1.SG  
'I am writing and reading it for Narmin'
- c. \* da=y-nus-im u da-xwên-im  
AM=3.SG-write.PRES-1.SG and AM-read.PRES-1.SG  
(putatively) 'I am writing it and reading it'
- d. da=y-nus-im u da=y-xwên-im  
AM=3.SG-write.PRES-1.SG and AM=3.SG-read.PRES-1.SG  
'I am writing it and reading it'

In (21-b), the clitic *-î* has large scope over the coordination of two verbs, which supports the phrasal affix analysis. In (21-c) however, where the clitic occurs within the word as an endoclititic, large scope over coordination is excluded and the clitic must be repeated on each conjunct, as is the case in (21-d). The same situation stands for the past transitive construction, in which the clitic marking subject-verb agreement has narrow scope over coordination of two verbs when it occurs as an endoclititic:

- (22) a. xwânû-kân=**m** kirî u frošt  
houses-DEF.PL=1.SG buy.PAST and sell.PAST  
'I bought and sold the houses'
- b. \* kirî=**m**-in u frošt-in  
buy.PAST=1.SG-3.PL and sell.PAST-3.PL  
(putatively) 'I bought and sold them'
- c. kirî=**m**-in u frošt=**im**-in  
buy.PAST=1.SG-3.PL and sell.PAST=1.SG-3.PL  
'I bought and sold them'

If clitics were regularly phrasal affixes, only one clitic would be expected in case of coordination of two verbs. The fact that the clitic does not have large scope in this case is again characteristic of lexical rather than phrasal affixes.

The facts examined in this section leads us to the following (temporary) conclusions: (a) the analysis of Sorani clitics as syntactic items is problematic (b) but a uniform phrasal affix analysis seems untenable also, since clitics exhibit a dual behavior. When occurring on the verbal head, they act very much like lexical affixes, otherwise they behave like phrasal affixes. These conclusions agree with those of Luis and Spencer (2004) for European Portuguese: endoclicis is considered as affixation to the verbal stem (i.e. lexical affixation), while proclisis as affixation to VP (i.e. phrasal affixation).

### 3. Absolute vs. simple prepositions

In addition to the facts considered up to now, the so-called ‘absolute prepositions’ and their clitic complements provide further significant evidence in favor of an affixal analysis of clitics in some contexts. As mentioned previously, contrary to personal endings, which exclusively occur with verbs, clitic pronouns may as well occur within a noun phrase, an adjectival phrase or a prepositional phrase, alternating with an independent pronoun or a noun phrase:

- |      |    |           |     |    |           |
|------|----|-----------|-----|----|-----------|
| (23) | a. | kitêb-e   | min | b. | kitêb-im  |
|      |    | book-EZ   | I   |    | book=1.SG |
|      |    | ‘my book’ |     |    | ‘my book’ |
| (24) | a. | bo        | to  | b. | bo=t      |
|      |    | for       | you |    | for=2.SG  |
|      |    | ‘for you’ |     |    | ‘for you’ |

Certain prepositions display two different forms depending on whether the complement they combine with is realized as a noun phrase or an independent pronoun, on the one hand, or a clitic pronoun, on the other hand. The first form is referred to as ‘simple’ while the second is called ‘absolute’. We will see that the latter class provides some crucial evidence in favor of the affixal analysis of clitic pronouns in Sorani Kurdish dialects.

Table (4) contains the list of simple and absolute prepositions of Suleymaniye Kurdish<sup>1</sup>. As one may notice, not all simple forms have a

<sup>1</sup>Apart from these two classes, there is a third class of prepositions including forms such as *sar* ‘top, head’, *pišt* ‘back, behind’, etc. These are initially nouns which have acquired a prepositional use. We will not deal with them in this paper.

corresponding absolute form and *vice versa*. Two prepositions, *bê* and *tâ*, never take a clitic complement, and the prepositions *bo* and *lagal*, which is a compound form, have only one form whatever the realization of their complement may be. The simple preposition *-a* and the related absolute form *-ê* are enclitics, which always attach to a verbal host.

| Sorani Kurdish prepositions |              |                |
|-----------------------------|--------------|----------------|
| Simple                      | Absolute     |                |
| <i>ba</i>                   | <i>pê</i>    | 'to, with, at' |
| <i>bê</i>                   | –            | 'without'      |
| <i>bo</i>                   | <i>bo</i>    | 'for'          |
| <i>-a</i>                   | <i>-ê</i>    | 'to'           |
| <i>la</i>                   | <i>lê</i>    | 'of, in'       |
| <i>tâ</i>                   | –            | 'until'        |
| –                           | <i>tê</i>    | 'to, with, at' |
| <i>lagal</i>                | <i>lagal</i> | 'with'         |

Table 4

Related simple and absolute forms generally differ as far as their prosodic properties are concerned: absolute forms bear stress, while *ba* and *la* are not stressed. The absolute preposition *-ê* is exceptional with this respect, in that like its simple counterpart, it is an enclitic. Furthermore simple forms undergo elision when followed by a complement beginning with a vowel:

- (25) b-am šaw-a  
 at-this night-DEF  
 'During this night'

At first sight, related absolute and simple forms seem to be allomorphs triggered respectively by the clitic versus non-clitic realization of the complement, as shown by the contrast between (a) and (b) in the following examples:

- (26) a. ba to/\*=t da-lê-m  
 to you/\*=2.SG AM-say.PRES-1.SG  
 'I am telling you'  
 b. pê=t/\*to da-lê-m  
 to=2.SG/\*you AM-say.PRES-1.SG  
 'I am telling you'
- (27) a. aw kuṛ-a zor l'êwa/=tân nâ-č-ê  
 this child-DEF very to you/=2.PL NEG-go.PRES-3.SG

- ‘This child does not resemble you very much’
- b. aw kuṛ-a zor lê=tân / êwa nâ-č-ê  
 this child-DEF very to=2.PL / you NEG-go.PRES-3.SG  
 ‘This child does not resemble you very much’

This is indeed what certain studies tell us about absolute prepositions: ‘These forms must be used when the preposition governs a pronoun expressed as an affix’ (Edmonds, 1955, p.496). However, this apparently clear picture of complementary distribution turns out to be misleading once other relevant data are taken into account.

First, when realized as a clitic, the complement of the preposition may be attached to a host different from the lexical head by which it is subcategorized for, and may come before or after it in the sentence. In examples (b) below, the clitic precedes the absolute preposition of which it is the complement.

- (28) a. **pê=y** bi-lê ka bi-rw-â  
 to=3.SG AM-tell.PRES that AM-go.PRES-??  
 ‘tell him/her to go’ (Edmonds, 1955, p.498)
- b. har wuša-yak=**y** **pê** a-lê-m  
 each word-INDEF=3.SG to AM-say.PRES-1.SG  
 ‘I shall say only one word to him’
- (29) a. **lê=yân** rojbâš a-kâ  
 to=3.PL good-morning AM-say.PRES  
 ‘He wishes them ‘Good Morning’ (Edmonds, 1955, p.498)
- b. rojbâš=**yân** **lê** a-kâ  
 good-morning=3.PL to AM-say.PRES  
 ‘He wishes them ‘Good Morning’

In this case, attachment to the subject NP seems possible:

- (30) Şêx Ahmad Pânke, Xwâ=**y** **lê** razî bê  
 Sheykh Ahmad Panke, God=3.SG of satisfied be.SUBJ  
 ‘Sheykh Ahmad Panke, I wish God be satisfied with him!’  
 (Bassols-Codina, 1992, p.150)  
 The following examples illustrates the attachment of the clitic after the preposition:

- (31) (Êwa) **pê=tân** wut-**im**  
 (You) to=2.PL tell.PAST-1.SG  
 ‘You told me’

Note that in this case, the absolute preposition can host a clitic with which it has no relation. In (31), for instance, the clitic attached to the absolute preposition realizes subject-verb agreement.

Second, absolute forms may be used in long distance dependency contexts, in relative clauses, for instance, where the ‘extracted’ complement of the preposition occurs in the matrix clause and has no local realization in the subordinate clause:

- (32) bird-iy-a                      aw sôn-a=y                      xoy    lê bu  
 take.PAST=3.SG=to DEM place-DEF=EZ<sup>2</sup> himself to be.PAST  
 ‘He took him to the place where he himself had been’

Note that in this case, a resumptive pronoun, realized as a clitic, usually occurs in the subordinate clause. Finally, and crucially, the absolute form is used for all cases where there is no overt syntactic realization corresponding to a complement, regardless of the fact that there is an underlying argument, as in (33), or no argument at all, as in (34). In these uses, which can be qualified as ‘intransitive uses’ the absolute preposition acts very much like an adverb:

- (33) komod-êk=im                      kirî                      u    kitâb-akân=im  
 cupboard-INDEF=1.SG buy.PAST and book-DEF.PL=1.SG  
**tê-dâ**<sup>3</sup>    dâna  
 to-POSTP put.PAST  
 ‘I bought a cupboard and put my books inside it’
- (34) gayîşt-im=ê  
 arrive.PAST-1.SG=to  
 ‘I arrived there’ (Edmonds, 1955, p.499)

Another intransitive use of absolute prepositions occurs in compound verb formation, illustrated by the following examples:

- (35) bo ci    pê=m    **pê** da-kan-î?  
 for what to=1.SG to AM-dig.PRES-2.SG  
 ‘Why are you laughing at me?’

In *pê kandin* ‘to laugh at’, the absolute preposition acts as a particle in conjunction with the verb and the whole sequence acts as a lexical unit, like phrasal verbs in English.

---

<sup>2</sup>The Ezafe enclitic, noted EZ, links the head noun to its modifier, here a relative clause.

<sup>3</sup>*dâ* is a postposition with a locative meaning .

On the basis of the facts examined in this section, the final picture for the distribution of simple and absolute prepositions can be revised in the following way:

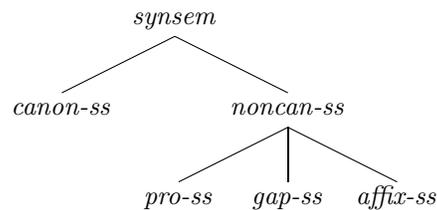
| Prepositions | Complement                                                                                                                     |
|--------------|--------------------------------------------------------------------------------------------------------------------------------|
| Simple       | Phrase (NP, PP, independent pronoun)                                                                                           |
| Absolute     | Clitic realized on the head<br>Clitic realized at distance<br>'Extracted' complement<br>No overt complement (intransitive use) |

Table 5

#### 4. A lexicalist account of simple/absolute alternation

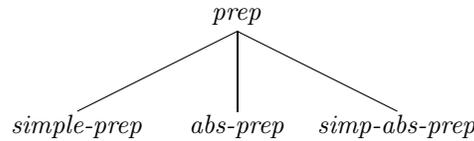
The fact that prepositions with a clitic complement are morphologically encoded like intransitive prepositions strongly suggests that clitics are not viewed as syntactic items. Furthermore, the fact that non-local (or 'gap') realization is treated on a par with clitic realization supports the lexicalist view of cliticization and extraction developed within HPSG (Bouma et al., 2001), where both pronominal cliticization (at least in some languages) and extraction are viewed as an instance of lexical alternation. The similarity between clitics and gaps is straightforwardly captured by the type hierarchy associated to these two types of signs: both clitics and gaps are a subtype of *non-canonical signs*. Furthermore, Sorani Kurdish seems to support the idea suggested by Miller and Sag (1997) that unexpressed complements may constitute a third subtype of *non-canonical signs*.

(36) *Synsem* type hierarchy



The (partial) type hierarchy associated with the prepositions is the following:

(37) (Partial) *preposition* type hierarchy



Under this view, the variation between simple and absolute forms of Sorani prepositions may be accounted for in terms of constraints on types of arguments associated with each class of preposition. Simple prepositions (*simple-prep*) take one canonical argument, while absolute prepositions (*abs-prep*) can only take one non-canonical argument. The argument structure associated to prepositions such as *bo* ‘to’, on the contrary, is not constrained and thus they can take either canonical or non-canonical arguments.

$$(38) \quad \textit{abs-prep} \Rightarrow [\text{ARG-ST } \textit{list}(\textit{non-canonical})]$$

$$(39) \quad \textit{simple-prep} \Rightarrow [\text{ARG-ST } \textit{list}(\textit{canonical})]$$

The following argument preservation rule constrains the way the preposition’s argument structure is realized :

$$(40) \quad \textit{Preposition Argument preservation}$$

$$\textit{prep} \Rightarrow \left[ \text{SS|CAT|LOC} \left[ \begin{array}{l} \text{SUBJ } \langle \rangle \\ \text{COMPS } \boxed{1} \\ \text{ARG-ST } \boxed{1} + \textit{list}(\textit{non-canonical}) \end{array} \right] \right]$$

This means that non-canonical arguments are optional in the COMPS list. In case the clitic is realized on the preposition, the latter projects a PP and has an empty COMPS list. The non-local realization of the clitic argument occurs only if the COMPS list contains a *non-canonical* member. We will deal with this case in the following section.

## 5. Non-local realization of the clitic complement

Once a lexicalist analysis is adopted, the question is how to deal with those cases where the clitic complement of the preposition is not attached to the proposition itself, but is realized at distance, occurring before or after it in the sentence, as in (29-b) and (31) above. At first sight, these examples seem to suggest that the clitic complement of an absolute preposition must be dealt with syntactically. It will be nonetheless argued that an affixal analysis is not only possible, but that it provides the most appropriate account of the data.

The first point to be mentioned is that the ‘non-local’ realization for

the ‘clitic’ complement of the absolute preposition is highly constrained and limited to two possibilities:

- (a) the ‘clitic’ argument is attached to the verb,
- (b) the ‘clitic’ argument is attached to the right edge of the constituent immediately preceding the preposition.

### 5.1 Verbal attachment

This attachment occurs only with transitive verbs in the past tenses, as shown by the contrast between the following sentences:

- (41) a. êwa **pê=tân** wut-**im**  
 You to=2.PL tell.PAST-1.SG  
 ‘You told me’
- b. \* êwa **pê** da-**m-lê-n**  
 You to AM=1.SG-tell.PRES-2.PL  
 (putatively) ‘You are telling me’
- c. \* êwa **pê** da-lê-**m-in**  
 You to AM-tell.PRES-1.SG-2.PL  
 (putatively) ‘You are telling me’

Crucially, the complement of the absolute preposition in this case is not realized as a ‘clitic’ but as a verbal personal ending. This constraint is not obvious in (41-a) above, since both the clitic and the personal ending for the first person singular are realized by the same form *-im*. The constraint becomes clear when other forms are taken into account:

- (42) a. rojbâš=yân            **lê** kird-**în**  
 good morning=3.PL to do.PAST-1.PL  
 ‘They wished us good morning’
- b. \* rozbâš=yân            **lê** kird-**mân**  
 good morning=3.PL to do.PAST=1.PL  
 (putatively) ‘They wished us good morning’
- (43) a. pâra-yêk-î            zor=î            **lê**    war  
 money-INDEF-EZ much=3.SG from back  
 girt-im  
 take.PAST-2.PL/3.PL  
 ‘He received a large sum of money from you/them’
- b. \* pâra-yêk-î            zor=î            **lê**    war  
 money-INDEF-EZ much=3.SG from back  
 girt=tân/yân  
 take.PAST=2.PL/3.PL

(putatively) ‘He received a large sum of money from you/them’

In case the direct object is also realized as a bound morpheme, the verbal stem bears two personal endings. The order in which the two affixes are placed seems to be subject to variation in different dialects and even within the same dialect. Edmonds (1955), for instance, claims that the affix corresponding to the complement of the absolute preposition precedes the affix corresponding to the direct object:

- (44) Xwâ **bo**=y nard-**im**-î(t)  
 God to=3.SG send.PAST-1.SG-2.SG  
 ‘God sent you to me’ (Edmonds, 1955, p.??)

Mackenzie (1961) gives the reverse order:

- (45) **lê**=y sand-in-**în**  
 for=3.SG take.PAST-3.PL-1.PL  
 ‘He took them for us’ (Mackenzie, 1961, p.116)

- (46) xwâ dâ=m-î-**n**=ê  
 God give.PAST=1.SG-3.SG-2.PL=to  
 ‘God gave me to you’

Note that in the last example, the verbal stem is followed by three bound morphemes: the personal ending *-m* refers to the direct object, the ‘clitic’ *-î* realizes subject-verb agreement and the personal ending *-in* is the argument of the absolute preposition *-ê*, an enclitic which is always attached to the verb.

These facts show that in the past transitive construction, the complement of an absolute preposition displays two different forms depending on whether it is attached to the preposition or to the verb. In the former case, it is realized by a ‘clitic’, while in the latter case, it is realized by a personal verbal affix:

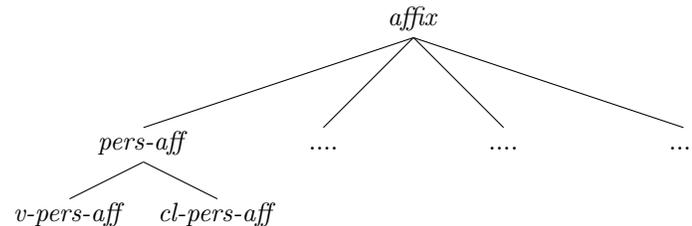
- (47) a. **pê**=mân=î dâ-n  
 to=1.PL=3.SG give.PAST-3.PL  
 ‘She/he gave them to us’  
 b. **pê**=y dâ-n-**în**  
 to=3.SG give.PAST-3.PL-1.PL  
 ‘She/he gave them to us’

This metamorphosis constitutes a serious problem for a syntactic analysis of the clitic placement. Indeed, a ‘clitic climbing’ account cannot

apply in this case, since the two occurrences of the ‘clitic’ are not identical. An affixal analysis, on the contrary, combined with an ‘argument composition’ operation, provides a straightforward account of the data under discussion.

Let us assume that *-mân* and *-în* in (47-a) and (47-b) above are both affixes. Thus, instead of having a class of clitics (i.e. syntactic items) and a class of personal endings (morphological items), we have a unified class of morphological objects, which will be referred to as *personal affixes*. These are a sub-type of affixes, and include in turn two sub-types: *verbal personal affixes* and *clitic personal affixes* :

(48) (Partial) affix type hierarchy



What distinguishes *verbal personal affixes* from *clitic personal affixes*, is that the former, being lexical affixes, are highly selective with regard to their host and can only be attached to a verbal stem. This attachment requirement can be warranted by a constraint on affixed words:

(49)  $v\text{-pers-aff-wd} \Rightarrow \text{word}[\text{HEAD } \textit{verb}]$

Since clitics are considered as affixes, clitic attachment is viewed as an instance of affixation. Let us return now to those cases where the ‘clitic’ argument of the preposition is realized on the verb. In this case the complement of the preposition is ‘reanalyzed’ as an argument of the verb. This analysis can be considered as an instance of ‘argument composition’, an idea developed in HPSG by Hinrichs and Nakazawa (1994) for German auxiliary constructions, and applied to ‘clitic movement’ in French by Miller and Sag (1997). Under this account, the verb, which is a functor, combines with an unsaturated argument – the absolute preposition – and inherits the ARG-ST requirements of the latter. The affixal complement of the absolute preposition is consequently realized as a verbal affixal argument. This operation is handled by a lexical rule, formulated as follows:

(50) Argument composition rule

$$\left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \textit{verb} \\ \textit{past} \end{array} \right] \\ \text{ARG-ST } \boxed{1} \left\langle \dots \text{PP}[\text{COMPS } \boxed{2} (\textit{affix})] \dots \right\rangle \end{array} \right] \rightarrow \textit{verb} [\text{ARG-ST } \boxed{1} + \boxed{2}]$$

This rule applies when a verb at the past tense that subcategorizes for a prepositional phrase combines with a preposition that has an element on its COMPS list. The affixal account for the ‘clitic’ can thus be maintained, since the ‘clitic’ becomes a verbal ending and is indeed a lexical affix.

## 5.2 The ‘clitic’ complement precedes the preposition

This attachment possibility, illustrated in (29-b) and repeated below, occurs either with intransitive verbs or with transitive verbs in the present tenses:

(51) rojbâš=yân            lê a-kâ  
 good-morning=3.PL to AM-say.PRES  
 ‘He wishes them ‘Good Morning’

Unlike verbal attachment, these cases challenge a strict lexicalist approach, since the ‘clitic’ attaches to a constituent with which it has no morphological or syntactic relation. Consequently one is inclined to assume that the preposition and the clitic should be conceived of as independent terminal nodes. There is however a crucial point to be mentioned: even though the ‘clitic’ is not attached to the preposition, a careful survey of data shows that it always occurs adjacent to the latter. Though I will not propose a formalized treatment for these cases in this paper, the lexicalist account can nevertheless be maintained. The ‘clitic’ is introduced in the lexical entry of the preposition, and thus the clitic and the preposition constitute a morphological unit of some sort. Let us further assume that preposition and the clitic are not strictly ordered and that the latter can precede or follow the former, though being an enclitic, the ‘clitic’ must attach to the left. A mismatch arises then when the ‘clitic’ is placed before the preposition: from a morphological point of view, the clitic goes with the preposition, though it is phonologically attached to a preceding element. In line with Kathol (2000), Crysmann (1999) and Crysmann (2003), I will assume that word-level signs can contribute more than one domain object into syntax. As noted by Crysmann (1999), such an approach weakens the

strict notion of lexical integrity, but it nevertheless enables us to flesh out the basic intuitions behind the lexicalist hypothesis.

## 6. Conclusion

Relying on the behavior of the so-called 'absolute' prepositions and their clitic complements in Sorani (Central) Kurdish, I have argued in this paper that the latter are best regarded as affixes, despite their apparent syntactic transparency. It has further shown that non-local attachment possibilities can be accounted for either in terms of argument composition with the verbal head of the sentence or in terms of linearization approaches. These two possibilities are in complementary distribution. The first one occurs exclusively in the past transitive construction, where the 'clitic' complement of the preposition is in some sort 'reanalyzed' as an Object complement of the verb and is consequently realized as a lexical affix following the verbal stem. The second attachment occurs either in the intransitive or in the present transitive constructions. Since the clitic complement in this case is adjacent to the preposition, it can be introduced as an affix in the lexical entry of the preposition, where the morpheme order is underspecified. Considered as an independent domain object, it can nevertheless be attached to a host different from the preposition, which accounts for the attachment to the element preceding the preposition.

## References

- Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.
- Bassols-Codina, S. 1992. *La Phrase Relative en Kurde Central*. Ph.D. thesis, Université de Paris 3.
- Blau, Joyce. 1980. *Manuel de Kurde*. Paris: Klincksieck.
- Bouma, Gosse, Robert Malouf, and Ivan A. Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19:1–65.
- Crysmann, Berthold. 1999. Morphosyntax paradoxa in fox. In *Constraints and Resources in Natural Language Syntax and Semantics*, pages 41–61. Stanford: CSLI Publications.
- Crysmann, Berthold. 2003. Clitics and coordination in linear structure. In *Clitics in Phonology, Morphology, and Syntax*, pages 121–159. Amsterdam: John Benjamins.
- Edmonds, C. J. 1955. Prepositions and personal affixes in southern kurdish. In *BSOAS XVII*, vol. 3, pages 490–502.
- Hinrichs, Erhrad and Tzuneo Nakazawa. 1994. Linearizing aux's in german verbal complexes. In *German in HPSG*, pages 11–37. Stanford: CSLI Publications.

- Kathol, Andreas. 2000. *Linear Syntax*. Oxford: Oxford University Press.
- Luis, Ana and Andrew Spencer. 2004. A paradigm function account of mesocclisis in european portuguese. In *Yearbook of Morphology*, pages 177–228. Dordrecht: Springer.
- Mackenzie, D. N. 1961. *Kurdish Dialect Studies*. Oxford: Oxford University Press.
- Miller, Philip H. and Ivan A. Sag. 1997. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 15:573–639.
- Zwicky, Arnold. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.



---

# Appalachian English *They*-Existentials

ERIKA TROSETH

## 1 Introduction

The ‘expletive’ in Appalachian English existential constructions can take different forms: *it*, as in (1), *there*, as in (2), and *they*, as in (3).<sup>1,2</sup>

- (1) It was a lot of them started bathing there. DOHP:JB
- (2) There was a spring out there. DOHP:JB
- (3) They was a lady around here. DOHP:CC

This paper investigates the variety of expletive construction in (3), analyzing these *they*-existential constructions as possessives.<sup>3</sup> More examples of this type are below.<sup>4</sup>

- (4) They’s a cemetery up there... DOHP:GAC

---

<sup>1</sup>I would like to thank Christina Tortora for discussing with me the data and ideas presented here and for very generously sharing with me the transcripts that she and Judy Bernstein produced while listening to recordings from the Dante Oral History Project housed at the Archives of Appalachia at East Tennessee State University. I would also like to thank Marcel den Dikken for discussing aspects of this paper with me. All errors and missteps are completely mine.

<sup>2</sup> Examples that come from the Dante Oral History Project will be indicated along with the initials of the speaker (Jerald Buttry:JB, Clyde Carter:CC, Gladys Amburgey Carter:GAC).

<sup>3</sup> Treating *they*-existentials as possessives was originally suggested to me by Christina Tortora.

<sup>4</sup> Examples that come from Montgomery & Hall (2004) will be indicated as M&H (2004).

- (5) They is not so many there now. M&H (2004:lxii)
- (6) They are another one down the street. M&H (2004:lxii)
- (7) They were several houses... DOHP:CC S

Sentences like the ones in (3) through (7) have been discussed elsewhere, including Tortora (forthcoming), Montgomery & Hall (2004), Wolfram & Christian (1976), and Hackenberg (1972), among others, and the interested reader is referred to these works.

This paper is organized as follows: Section two of this paper is dedicated to the nature of the expletive in sentences like the ones in (3) through (7). Section three provides background to and develops the possessive analysis pursued here. Section four addresses the apparent flexibility *they*-existentials exhibit with respect to verbal morphology and the nature of the ‘associate’. This apparent flexibility is seen in (3) through (7) above, in which it *appears* to be the case that plural and singular ‘associates’ can optionally appear with a singular verb form (-s) or a plural verb form (-r). For much of this paper, forms of *be* will be described as -r forms or -s forms rather than plural or singular for reasons that I hope will become clear in section four. Section five concludes the paper.

A few terms will be used throughout this paper, and it is worth dedicating some space to them. I will be referring to ‘Appalachian English’ although I do not truly take it to be the case that only one dialect of English is spoken in Appalachia. The construction discussed here is often called an existential construction, and the preverbal lexical item in these constructions is often called an expletive. I will use the term ‘expletive’ without intending to take a stand one way or the other on the true nature of so-called expletives. The postverbal NP in these sentences is often called the ‘associate.’ Because I treat these sentences as possessive constructions, I do not view the postverbal NP in this way. Finally, when I wish to distinguish between the *they* that is used in Appalachian English existential constructions (*They is a man in my room*) and the *they* that is used in other constructions (*They are planning to arrive at seven*), I will describe the latter *they* as ‘referential’ *they*, without intending to make specific claims about reference.

## 2 Two *Theys*

Above it was shown that the apparent agreement marking on the verb in *they*-existentials does not necessarily correspond to the apparently plural Nominative expletive nor to the ‘associate’, be it singular or plural. Often, speakers who allow -s verb forms in *they*-existentials do not exhibit this

phenomenon in their referential uses of *they*. Why should this be? The proposal advanced in Tortora (forthcoming) that Appalachian English has two *theys* is useful in addressing this question and is adopted here.

### 2.1 Weak *They* and Strong *They*

A number of tests have been employed to distinguish strong locatives from weak locatives (Allan 1971; Tortora 1997; Cresti & Tortora 2000) and strong pronouns from weak pronouns (Cardinaletti & Starke 1999), including coordination, modification, clefting, and the ability of the pronoun or locative to bear stress. The weak versions of these items are typically taken to be impoverished in some way compared with the strong version. The *they* in Appalachian English *they*-existentials is taken here to be impoverished with respect to its number value, following the proposal in Tortora (forthcoming) that referential *they* (strong *they*) has the number specification [-singular] while the *they* in *they*-existentials (weak *they*) is unspecified for the value of this feature and is marked [singular] (which is distinct from the specification [+singular]).

### 2.2 Weak *They* as a Nonreferential Possessor

The *they* in Appalachian English *they*-existentials is here characterized as a nonreferential possessor. This is similar to the *they* in the sentence below, which can be paraphrased as ‘There are lots of distance runners in Oregon.’

- (8) They have lots of distance runners in Oregon.<sup>5</sup>

### 2.3 The Failure of the ‘Associate’ to Raise

Many analyses of *there*-existentials assume that agreement between the verb and postverbal NP (the ‘associate’) obtains through the raising at LF of the associate or features of the associate. Under the analysis of *they*-existentials proposed here, the postverbal NP does not raise at LF just as the postverbal NP does not raise at LF in possessive sentences that have strong pronouns or full DPs as the possessor. The postverbal NP is not the associate in *they*-existentials, but rather is the possessed item.<sup>6</sup>

The view on the pronominal subject of *they*-existentials has been covered, and the next section addresses the form of the verb in *they*-existentials, which is *be*, not *have*.

---

<sup>5</sup>Thanks to Christina Tortora for bringing examples like these to my attention.

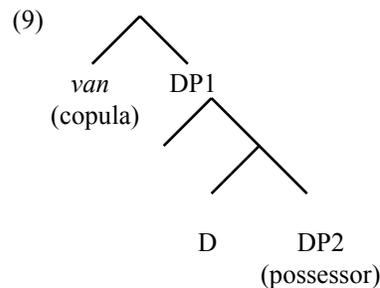
<sup>6</sup>For a different view on the absence of ‘associate’ raising in *they*-existentials see Tortora (forthcoming).

### 3 Kayne (1993)

In this section I will briefly present the details of Kayne (1993) that are relevant to the proposal advanced here. Some modifications will be made to the picture that Kayne presents. The end proposal is that *they*-existentials in Appalachian English be analyzed as underlyingly possessive constructions.

#### 3.1 Hungarian and English Possessives

Kayne (1993) takes the analysis of Hungarian possessives presented in Szabolcsi (1981, 1983) as a starting point and develops an account of possessive *have* that he generalizes to auxiliary *have*.<sup>7</sup> The basic components of Szabolcsi's analysis as Kayne presents them follow. The copula selects a DP headed by a D that itself selects a possessor DP. This is depicted below.

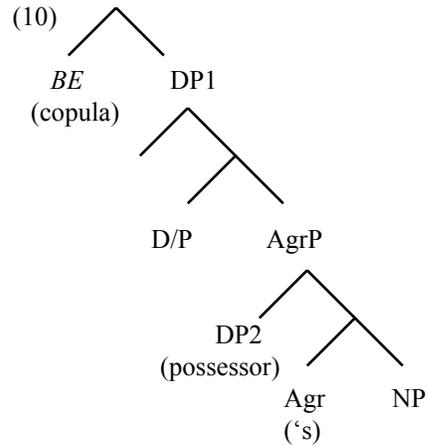


There are three possibilities regarding DP2: DP2 remains in situ and is marked Nominative; DP2 raises to SpecDP1 and is marked Dative; DP2 raises to SpecDP1, is marked Dative, then raises to the left of the copula.

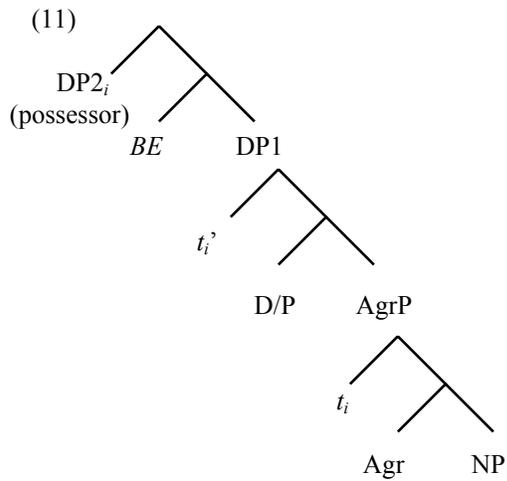
Kayne's proposal for English possessives is similar in that the copula selects a DP. A modification he makes is that D selects an Agreement Phrase instead of a DP possessor. The possessor occupies SpecAgrP, and the possessed NP is the complement of the Agr head. Accepting the parallel between CP and DP argued for by Szabolcsi, Kayne proposes that the head of DP1 can be a D or a P (on analogy with prepositional complementizers found in the head of CP). This is illustrated below.

---

<sup>7</sup> For problems with such an account, the interested reader is referred to Tham (2004). Thanks to Steve Wechsler for this reference.



Kayne suggests that it is reasonable to posit 's as an instantiation of the Agr head if this 's is taken to be similar to the -s ending on third person singular verbs. Having established an underlying structure, Kayne first suggests the derivation represented below.



The possessor raises to SpecDP1, as in the Hungarian case. English then must do what Hungarian has the option of doing: The possessed DP raises

further, to the left of the copula. If nothing more were said, this would yield sentences like *Lee Morgan is a new album* that would be interpreted as *Lee Morgan has a new album*. Kayne considers dialects of English in which this is impossible, and so he modifies the proposal to rule out a *be* possessive and to guarantee a *have* possessive.<sup>8</sup>

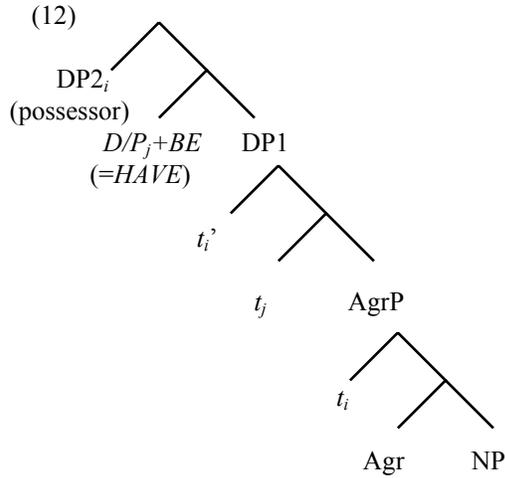
However, in the Appalachian English examples under consideration here, the proposal is precisely that in *they*-existentials we see a possessive construction with a form of the copula rather than the verb *have*. Although the further modifications Kayne makes to rule out the presence of the copula in English possessive constructions will not be adopted in this analysis of *they*-existentials, some of the details will be relevant, so I review them here briefly.

### 3.2 A- and A-bar Positions

Kayne suggests that a modification of the derivation represented in (11) is forced once the nature of specifier positions is taken into consideration. Assuming the parallel between CP and DP, Kayne proposes that SpecDP1 is an A-bar position. If this is true, in (11) above the possessor DP moves from an A-position to an A-bar position and then to an A-position, the final step being an instance of improper movement. If this sequence of movements is to be made licit, SpecDP1 must be made into an A-position or all subsequent movements by the possessor NP (after it has landed in SpecDP1) must be to A-bar positions. Kayne proposes that if the D/P head of DP1 incorporates into the head that hosts *BE*, SpecDP1 will become an A-position. Building on the idea (Freeze 1992) that *HAVE* = *BE* + *P*, Kayne proposes that when the D/P head incorporates into *BE* the complex head is pronounced *have*. This modified derivation is illustrated below.

---

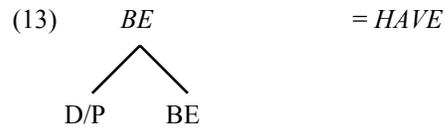
<sup>8</sup> He does not in fact explicitly state which dialect of English is being considered.

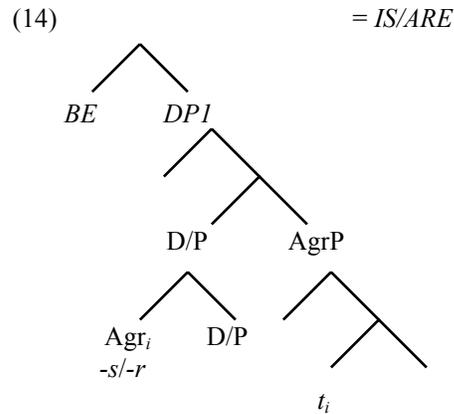


### 3.3 Incorporating Agr

The first modification I would like to make to the above approach in order to account for Appalachian English *they*-existentials is that instead of D/P incorporating into *BE*, the Agr head incorporates into D/P. Although this is a modification to the picture Kayne presents for English, it is in fact identical to the proposal he makes for Central Italian dialects for cases in which *be* is the auxiliary form rather than *have*. A second modification I make is that the Agr head may sometimes host an *-s* and sometimes host an *-r*. This will be discussed briefly in section 4.

When the phonologically null complex D/P head incorporates into *BE* the resulting surface form is *have*, as Kayne suggests. However, if instead the Agr head incorporates into D/P and no further incorporation obtains, the result is *is* or *are*, similar to what he suggests for the relevant cases in the Central Italian dialects. These possibilities are illustrated below.





The derivation represented in (14) yields *is* or *are* depending on the nature of the Agr head. It is open at this point whether the *-s* and *-r* correspond to a specific value associated with number agreement. My view is that no matter the surface form of the verb (*have*, *is*, or *are*) the derivations represented in (13) and (14) are associated with a possessive interpretation that comes about because of their identical underlying structure.

The nature of SpecDP1 is worth considering here. In the English examples Kayne originally considers, this is taken to be an A-bar position unless the D/P head incorporates into *BE*. This incorporation yields *have* as the surface verb form. Kayne proposes that incorporation of the Agr head into the D/P head can also transform SpecDP1 into an A-position when this incorporation obtains. Another possibility to consider is that SpecDP1 is simply not projected in Appalachian English. The position is needed in the Hungarian case—it is the site in which Dative case is assigned. But in the English examples it seems that this position is superfluous as case (Nominative) is assigned above this position.

## 4 Agreement

### 4.1 Default Agreement and Actual Agreement

The *-s* ending on the verb in *they*-existentials could indicate singular agreement or default agreement. Tortora (forthcoming) argues that expletive *they* triggers default agreement. As mentioned above, she argues that expletive *they* has a NumP projected in its syntactic structure, but does not have a value specified—it is [singular] in contrast with [-singular] (which is the case for strong/referential *they*) and [+singular] (which is the case for singular pronouns like *she*). The suggestion made above that Agr in *they*-

existentials can host an *-s* is compatible with this approach as long as it is taken to be default rather than singular agreement. The suggestion that Agr in *they*-existentials can host an *-r* may also be compatible with this approach, but this requires a closer look at the data recorded for individual speakers. Some discussion of this follows.

#### 4.2 An Apparent Free-For-All

The following examples might lead one to believe that in *they*-existentials agreement is completely optional in the sense that any verb form (*-s* or *-r*) can be used with any NP type (singular or plural).

- |                                            |                 |
|--------------------------------------------|-----------------|
| (15) They's a cemetery up there...         | DOHP:GAC        |
| (16) They is not so many there now.        | M&H (2004:lxii) |
| (17) They are another one down the street. | M&H (2004:lxii) |
| (18) They were several houses built...     | DOHP:CC         |

However, taking into consideration facts that have been reported regarding the form of singular and plural marked verbs and the sentences produced by individual speakers, there is nothing in the data above to necessarily suggest that *any individual grammar* instantiates optionality. There is also nothing in the data above, taken alone, to determine the status (singular, default, or plural) of the *-r* verb forms. Examples in which *are* and *were* appear with singular referential subjects can be found in Montgomery & Hall (2004) and discussion of these forms is found in Tortora (forthcoming). What is important here is simply that the reader not assume that verbal forms that may be necessarily singular or plural in her own dialect are necessarily singular or plural for the speakers who produced the sentences above.

The sentence below is worth considering together with (18).

- (19) They was two men would walk these logs... DOHP:CC

Example (18) is unusual in that the past tense form of the verb *be* is *were*. In many Appalachian dialects the past tense form of *be* is *was* across NP types. These two examples—since they were produced by the same speaker—raise questions about the possibility that speakers may shift between two grammars. Defining the precise nature of the two grammars would surely be a difficult (though rewarding) challenge, and it is not something that can be done on the basis of this transcript. These examples also raise questions about the possibility of optionality within a grammar, which, depending on one's theoretical framework, will be faced differently. Finally, these examples illustrate in part the limits of linguistic analysis that

can be seriously pursued without fieldwork that would accompany and supplement existing recordings and writings. The example below further illustrates this point. Here the speaker corrects the form of the verb in the sentence (this is a *there*-existential rather than a *they*-existential).

(20) There certainly were—was—witches.<sup>9</sup>

Had this speaker not made this correction it might appear that *were* was the intended form and the assumption might be made that the speaker would accept as grammatical the sentence above with *were* rather than *was*. A precise proposal regarding the nature of Agr (as discussed here) and the nature of *-s* and *-r* verb forms would require more complete knowledge of the grammar of *individual speakers*.

## 5 Conclusion

It has been argued that *they*-existentials in Appalachian English are underlyingly possessives. Following the approach to possessives outlined in Kayne (1993) and Szabolcsi (1981, 1983) a proposal has been advanced that accommodates a weak nonreferential possessor *they* and a surface verb form of *be* rather than *have*. Some discussion has been offered regarding the nature of the Agr head that is a piece of this proposal and the challenges in interpreting *-s* and *-r* verbal forms in existential and other constructions.

---

<sup>9</sup>This example is from the Joseph S. Hall collection and is included in the transcriptions of tapes provided to me by Christina Tortora.

## References

- Allan, K. 1971. A Note on the Source of *There* in Existential Sentences. *Foundations of Language* 7:1-18.
- Cardinaletti, A. and M. Starke. 1999. The Typology of Structural Deficiency. *Clitics in the Languages of Europe*, ed. H. van Riemsdijk, 145-233. Berlin: Mouton de Gruyter.
- Cresti, D. and C. Tortora. 2000. Aspects of Locative Doubling in Resultative Predication. *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society*, eds. S. S. Chang, L. Liaw, and J. Ruppenhofer, 62-73.
- Freeze, R. 1992. Existentials and Other Locatives. *Language* 68:553-595.
- Hackenberg, R. 1972. A Sociolinguistic Description of Appalachian English. Doctoral dissertation, Georgetown University.
- Kayne, R. S. 1993. Toward a Modular Theory of Auxiliary Selection. *Studia Linguistica* 47:3-31.
- Montgomery, M. and J. Hall. 2004. *A Dictionary of Smoky Mountain English*. Knoxville: University of Tennessee Press.
- Szabolcsi, A. 1981. The Possessive Construction in Hungarian: A Configurational Category in a Non-Configurational Language. *Acta Linguistica Academiae Scientiarum Hungaricae* 31:261-289.
- Szabolcsi, A. 1983. The Possessor That Ran Away From Home. *The Linguistic Review* 3:89-102.
- Tham, S.-W. 2004. Representing Possessive Predication: Semantic Dimensions and Pragmatic Bases. Doctoral dissertation, Stanford University.
- Tortora, C. 1997. The Syntax and Semantics of the Weak Locative. Doctoral dissertation, University of Delaware.
- Tortora, C. forthcoming. The Case of Appalachian Expletive *they*. *American Speech*.
- Wolfram, W. and D. Christian. 1976. *Appalachian Speech*. Arlington: Center for Applied Linguistics.



---

# Converb Constructions in Darma— A Tibeto-Burman Language

CHRISTINA M. WILLIS

## 1 Introduction

Cross-linguistically languages are found to mark verbs in dependent clauses differently than verbs in matrix clauses. There are special verb forms that mark relative clauses, complement clauses, and adverbial clauses. In this paper, I will examine a morpheme that marks an adverbial clause in Darma a Tibeto-Burman (TB) language spoken in India.

Chaining clauses using nonfinite verbal markers is an areal feature of South Asian Languages (Masica 2005). These markers have several terms associated with them. For example, in Hindi pedagogical grammars, the clause-chaining marker is called a ‘conjunctive particle’ (Snell 1992: 128), while some linguists call it a converb (Haspelmath 1995: 2; 2002: 67):

(1)    **banie ke        bete ne        ciŋ<sup>h</sup>ii        lik<sup>h</sup>-kar**  
         grocer POSS    son    ERG letter(F).SG    write-CONV

**ɕaak mē ɕaal-ii**  
box    in    put.PAST-F.SG

*The grocer’s son wrote and posted a letter.*

*(LIT: ‘having written a letter, posted (it).’)*

(Haspelmath 1995: 2; 2002: 67)

*Texas Linguistics Society 9: Morphosyntax of Underrepresented Languages*

Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, and Jessica White (vol. eds.) and Stephen Wechsler (series ed.).

Copyright © 2007, CSLI Publications.

These constructions are well documented in the Indo-Aryan and Dravidian languages of the subcontinent (Masica 2005). While there are many TB languages in South Asia, many are not well documented, so it is difficult to determine whether they share this areal feature. The focus of this paper is an underdocumented TB language, Darma, which shares the areal feature of linking verbs and clauses using ‘converbs’ with its Indo-Aryan and Dravidian neighbors.

The Darma data presented in this paper were collected during three separate fieldtrips to Dharchula, India<sup>1</sup>. I am in the process of analyzing the data gathered, in an effort to complete a descriptive grammar of the language. All analyses presented here are representative of a work in progress.

This paper is organized as follows: The first section will provide a brief sketch of the Darma community and the linguistic situation of the area, and a brief overview of the structure of the language; section two introduces the term ‘converb’ and presents a working definition for the discussion; in section three, I will outline data from Darma that supports the hypothesis that there is a converb in the language; section four will examine data that does not fit within the parameters of a ‘converb’; in section five, I will introduce the term ‘switch reference’ and reexamine the Darma data with this analysis. In section six, I will examine data that does not fit the definition of switch reference as presented in Longacre (1985) and Haspelmath (1995). In sections seven and eight, I will reconsider switch reference using a reanalysis of the definition from Mithun (1993). Finally, I will conclude with a summary of my findings and suggest future research.

## 2 The Darma Language and its People

Darma<sup>2</sup> is a Tibeto-Burman language spoken in the area of the Himalayan town of Dharchula, which is on the border of Nepal, close to Tibet. Darma is a verb final language, with postpositions. Agents are case marked with an ergative morpheme, although this morpheme is sometimes omitted following the pattern in Hindi, where ergative case is marked only in the simple past.

---

<sup>1</sup> Research was supported by a Liberal Arts Graduate Research Award from the University of Texas at Austin, Fulbright IIE, Fulbright-Hays, and the National Science Foundation (BCS 0236475; Anthony C. Woodbury, PI).

<sup>2</sup> Darma is sometimes called Darmiya in the literature. It is also sometimes lumped together with its sister languages Byansi and Chaudangsi under the term Rangboli. While only preliminary work has been done with all three varieties, it is believed that Darma is distinct from Byansi and Chaudangsi (Gordon 2005; Trivedi 1991; Krishan 2001).

According to the Government of India's 2001 census data, the population of Darma people<sup>3</sup> is approximately 2,600. Based on my personal experience within the Darma speaking community, not all of the people who identify themselves as Darma speak the traditional language. The local varieties of Nepali and Kumauni are spoken by most locals, and Hindi is taught in the schools. I have met many school age children who speak Hindi, but do not understand or speak Darma. The children who do speak Darma fluently and who use it as a daily form of communication are often from families that are not financially secure.

Historically, the Rang people were the dominant population in the Dharchula area. After China invaded Tibet, and almost took over the region of India where the Rang people live in 1962, a large military presence was established in the Dharchula area. Additionally, in the last five years the government has been working on a small dam in the lower regions of the Darma Valley. These circumstances have brought many Hindi speaking people from the plains to this mountain community. I was told that this influx of outsiders has altered the community both culturally and linguistically.

The main consultant<sup>4</sup> throughout my research project was Bishan Singh Selal, a 25 year old male who was educated in both English and Hindi. He has a BA and is working on a Masters of Education. He teaches at the local army outpost elementary school, where he instructs in English. His mother speaks minimal Hindi, which is one reason why Darma is still the primary language of the household. Based on interviews and personal interactions with Darma speakers, I have found that the majority of monolingual Darma speakers in the community are those women who received no formal education.

---

<sup>3</sup> The government of India classifies the Darma people, along with their sister clans the Byansi and Chaudangsi, under the term Bhotiya. Darma is a self-referent, while the term Bhotiya is resented by the people. The self-referent for the Darma, Byansi, and Chaudangsi as a group is Rang.

<sup>4</sup> I would like to thank the Darma people who graciously helped me gather data, and allowed me to record conversations, songs, personal narratives, and stories. So many people contributed to my project by answering my endless stream of questions regarding the language and culture. There are too many names to list here, but the kind people in Gothi and Kalika must be recognized. I would especially like to thank my primary consultant Bishan Singh Selal, who worked with me in addition to working another job fulltime.

### 3 Converbial Constructions Defined

The first issue that must be addressed is how I will classify the structures under investigation. In Darma it is possible to have a sentence with multiple verbs as illustrated in (2) below.<sup>5</sup>

- (2)    **hã    ʈi-lən        jo    hre-nu        ni-n-cu.**  
           then lead-CONV down bring-ADJR be-1pl-PST  
           *Then, we brought (the yak) down leading (it).*                      T0018:020

There has been much debate over what a converb construction is (Haspelmath 1995), in part because the term ‘converb’ is not used consistently in the literature; what would be called a ‘converb’ by one scholar may be called a ‘gerund’, ‘conjunctive participle’, or ‘adverbial participle’ by another in the literature.<sup>6</sup>

In an attempt to clarify the question of converbs, Haspelmath and Koenig (1995) present a collection of papers wherein a formal definition of converbs is proposed, and the distribution of the morpheme is explored cross-linguistically. In the introductory chapter, Haspelmath defines a converb as ‘a nonfinite verb form whose main function is to mark adverbial subordination.’ Converbs lack tense, aspect, mood and agreement markers, and are thus nonfinite (1995: 3-4). The typical structure of a converb is STEM + AFFIX and they are commonly found in verb final languages (Longacre 1985: 264; Haspelmath 1995: 9).

Haspelmath (1995) discusses the fact that converbal constructions have been analyzed in varying ways in the literature. In addition to different terms to describe the same phenomenon, there are other constructions that

---

<sup>5</sup> The data presented in this paper are in the IPA. The format of presentation is three lines: the top line is the phonemic transcription with morpheme boundaries; the second line contains the morpheme gloss; and the third line is the free translation. The numbered code that follows is a reference to the source text for the data. Some of the data are not from natural discourse. I will indicate that those data are from elicitation sessions. The abbreviations are as follows: ADJR = adjectivalizer; ASP = aspect; CONV = converb; ECHO = echo formation; EMPH = emphasis; ERG = ergative; FOC = focus; FUT = future; GEN = genitive; (H) = a Hindi word; INF = infinitive; INTJ = interjection; LOC = locative; NEG = negative; PL = plural; POSS = possessive; PRS = present; PST = past; number is marked 1sg, 1pl, 2sg, 2pl, 3sg, 3pl. Terms such as ASP and PST may change after further analysis.

<sup>6</sup> Longacre (1985) refers to ‘clause chains’ to describe what Haspelmath calls a converbal construction as in example (1). Haspelmath (1995: 45-46) explicates which term might be used where: The term ‘conjunctive participle’ is used in descriptions of South Asian languages; the term ‘gerund’ is used in descriptions of Romance languages; and the term ‘adverbial participle’ is used in descriptions of Slavic languages.

are similar to converbs, but don't meet all of the criteria Haspelmath outlines in his definition of the term. These other constructions include:

- copredicative participles;
- medial verbs;
- clause chaining;
- absolute constructions;
- infinitival constructions.

The structure presented in (1) is identified as a converb by Haspelmath (1995: 2; 2002: 67). According to Longacre (1985), the structure in (1) would be classified as a 'clause chain'. Haspelmath (1995: 8) claims that 'clause chaining' is a structure that is very similar to the converbal construction, but the two are not the same. What differentiates the two is that converbs are subordinate clauses, while clause chained constructions are not.

Longacre states that the structures in clause chains consist of 'sentence margins' and 'sentence nuclei'. The verbal element in a sentence margin is termed a medial verb and is usually marked with morphology that indicates its status as a medial verb. Medial verbs also have morphology that indicates a switch of reference, which will be discussed further below. The 'sentence margins' in clause chains are subordinate to the 'sentence nuclei', while the 'sentence nuclei' are independent clauses and contain the main meaning of the utterance. The verbs in sentence nuclei are more complete than the medial verbs, meaning the verbs in the nuclei contain all of the tense, aspect, and agreement morphology required by the syntax.

According to Longacre, another feature that distinguishes clause chains from similar formations like conjoined clauses is that temporal overlap and temporal succession are features of clause chains (1985: 264). In summary, both Haspelmath and Longacre suggest that these constructions are part of a subordinate clause. Haspelmath requires that the converb bear no tense, aspect, or agreement morphology, and Longacre requires that the clause chain bear a morpheme identifying the subject of the following clause as being the same or different from the preceding clause.

Clause chains as defined by Longacre and converbs as defined by Haspelmath are so similar that the dueling definitions serve as an example of the confusion that multiple terms for the same phenomenon can cause. Both constructions have a nonfinite verb in a subordinate clause, which is associated with a matrix clause that contains a fully inflected verb.

For this paper, I will follow the definition of converbs outlined by Haspelmath, in part because it the most recent source. Haspelmath states that the difference between a clause chain and a converb is that converbs are subordinate clauses, while clause chains are not. While Longacre claims

that clause chains are subordinate clauses, he does not demonstrate subordination with any tests. Haspelmath, on the other hand, proposes some tests to determine whether the construction in question is in fact a subordinate clause or not. Per his definition, only those constructions that are subordinate can be classified as true converbs.

The decision to follow Haspelmath's definition is rather arbitrary since his definition is so close to Longacre's, but Haspelmath and Koenig have attempted to unify the theory of these constructions and do away with multiple terms for the same phenomena, and I think this is the right step toward both furthering our understanding of converbs cross-linguistically and developing a unifying theory of this phenomenon. I will, however, revisit Longacre's definition later in the discussion and suggest that some components of his definition need to be incorporated into Haspelmath's account of this cross-linguistic phenomenon. I will use examples from Darma to support Genetti's suggestion that the term 'converb' needs to be revised (2005: 81).

#### 4 Converbs in Darma: [-lən]

I will consider the structures found in Darma, like example (2) above, to be converbs as defined by Haspelmath (1995). The converb<sup>7</sup> in Darma is an affix on a verb stem: V<sub>STEM</sub> + [-lən], as is the case cross linguistically. In this section, I present data that supports this claim.

According to the working definition for this paper, converbal constructions are subordinate clauses. Haspelmath (1995) states that subordinate clauses can:

- embed in the superordinate clause;
- come before or after the superordinate clause;
- be focused with 'only' and 'also' particles;
- be extracted; meaning it is possible to form a question based on a subordinated clause.

Additionally, in subordinate clauses:

- there is backward pronominal anaphora and control.

Converbal constructions in Darma are found primarily in narratives. This is logical considering the purpose of this type of construction is to move the narrative forward (Saxena 2004). Because most of the converb

---

<sup>7</sup> There is another suffix found in Darma, [-tʰe], which I have also identified as a converb. The [-tʰe] is not used as frequently as [-lən], but it is used in a similar fashion. My consultant states that the two are always interchangeable, but further investigation is needed to make this claim. I will not include [-tʰe] in the discussion here.

data I have in Darma come from natural discourse, I do not have examples of sentences for each of Haspelmath's tests outlined above. In fact, during elicitation sessions, manipulating converbal constructions was a frustrating task for me and my consultant. I will address this issue later in the paper.

The data that I have for Darma suggest that the constructions in question are true converbal constructions; meaning they are subordinate clauses. Of the five tests proposed by Haspelmath, I have data to support two tests: Darma converbal clauses can be placed before and after the superordinate, or matrix, clause; and Darma converbs can take a focus particle. This is shown in examples (4-7). In example (3), we see a structure similar to the one presented in (2) above. The converb precedes the matrix verb in the utterance. Here we have an echo formation, and each component of the echo<sup>8</sup> bears the converb morpheme.

- (3) **hã bakca t̪ʰuŋ-lən t̪ʰaŋ-lən de-çjen,**  
 then wedding dance-CONV ECHO-CONV go-1pl.PRS  
**iɔ̃ su.**  
 after\_that  
*Then, we go to the wedding dancing and everything after that.*  
 T0013:013

In examples (4-5), we find the converb in a sentence final position, where it has been extraposed. Darma is a verb-final language, and as we saw in examples (2-3), the converb usually precedes the fully inflected verb (i.e. the verb with tense, aspect, and agreement morphology).

- (4) **mi-çən kʰe mə-ra-yəŋ wi-lən.**  
 person-PL some NEG-come-FUT invite-CONV  
*Some of the people will not come after being invited.* T0037: 019
- (5) **ɟjadu le ɟa-n ni-n-ɟu darma ra-ləŋ.**  
 a\_dish.food also eat-ADJR be-1pl.PST Darma come-CONV  
*We also used to eat flour porridge after coming to Darma (Valley).*  
 T0021:015

The examples in (6) and (7) have the converb with a focus particle, [-na],

<sup>8</sup> Echo formations are an areal feature of South Asian languages. A word is paired with a rhyming word, and contributes a meaning 'and everything that goes along with that'. For example in Hindi 'chai vay' means 'tea and everything that goes with that', like biscuits and conversation. Here 'chai' means 'tea', and the second word 'vay' holds no meaning on its own. In Darma both the stem that has meaning and the echo element are marked as converbs.

which is usually translated as ‘also’ or ‘only’.

- (6) **jil-lən-na**            **hre-mu.**  
 rub- CONV-FOC    bring-INF  
*You should bring (it) while rubbing only.*    T0042:elicited sentences:237
- (7) **gəni**            **b<sup>h</sup>arət**            **k<sup>h</sup>əxtu**            **nagrikta**  
 2pl                  India                  from                  citizenship  
**k<sup>h</sup>ət-lən-na**            **hre-ni, or ɸo ju nagrikta**  
 sever-CONV-FOC    bring-2sg and here of citizenship  
**joynd**            **gani.**  
 join                  do-2sg  
*"You should just sever your citizenship with India and then come here and become a citizen."*                  T0025:066

These data demonstrate that the clause with the converb in Darma is indeed subordinate per the tests proposed by Haspelmath (1995); based on these data, I will conclude that Darma does have converbal constructions.

## 5 Contradictory Data

Considering that part of the definition of converbs is that they are nonfinite constructions, it is puzzling that I find data as in (8) and (9) with the morpheme [-ju]<sup>9</sup> that marks past tense in Darma.<sup>10</sup>

- (8) (i) **ʔo**    **esa**                  **he**    **jo**    **ʔu**    **su**    **jo ni-ni**  
 then of\_this\_type.(H) be.(H) that 3sg ERG um  
 (ii) **k<sup>h</sup>rapat**            **ɸimag** **ga-lən-ju,**            **ʔu**    **su**  
 mischievous\_idea mind do-CONV-PST 3sg ERG  
 (iii) **ga<sup>h</sup>o**            **me**    **po-lən-ju,**            **pulis**    **su**    **jo nini,**  
 flour\_mill fire start-CONV-PST police ERG INTJ

<sup>9</sup> The morpheme [-ju] alternates with [-cu] when attached to a verb stem (see Table 1 in Section 6 below) and with [-cu] and [-jaŋ] when attached to a converb.

<sup>10</sup> Thank you to Carlota Smith for pointing out the problem of introducing this morpheme [-ju] as a past tense marker. I am not certain that it is in fact tense, but until I have an analysis that I find satisfactory, I will call it past tense. The utterances that contain this morpheme are consistently translated to past in both Hindi and English. I revisit this issue later in the paper.

(iv) **tum-lən-juu**      **jo nini,**    **raja**    **ḍərosar**  
 catch-CONV-PST    INTJ    king    near    deliver

(v) **pu-kur-su.**  
 ASP-take\_away-3sg.PST  
*So it is like this, after he thought the mischievous thoughts, after he started the mill fire, after the police, um, caught him, uh, they delivered him to the king.*  
 T0025:019-020

(9)    **ji**    **pe-lən-juu**                      **geje-o.**  
 1sg    slip-CONV-PST    fall\_person-1sg.PST  
*I slipped and fell.*  
 T0042 Elicited:367

Considering the meaning of (8) and (9), it appears that the past tense marker on a converb construction imparts a sense of completion and should be interpreted to mean ‘after’. An analysis such as this would actually suggest that the morpheme meaning ‘after’ has been grammaticalized as the past tense marker. We can see that interpreting [-juu] as ‘after’ would make sense in an example like (8) where a series of events are occurring in a sequential order. First the hero gets a mischievous idea in his head, and after that he burns down the mill, and after that he is caught by the police.

But then it is necessary to consider examples like (10-11) where there is no past tense marker [-juu] on the converb, and each sentence still has a sequential interpretation.

(10)    **hã niŋ ḍa**      **jo nini re**    **ru juu ja**  
 yes 1pl EMPH    that\_is field in from tuber\_wild  
  
**kwe-lən**      **ja-nu**      **ni-n-juu.**  
 boil-CONV    eat-ADJR    be-1pl-PST  
*Yes, we even, that is, we ate the wild yams from the field after boiling them.*  
 T0021:003

(11)    **jjadu**      **vol**    **tsjokpa**      **ga-lən**  
 a\_dish\_food greens veg\_cooked    make-CONV  
**jan**                      **ni-n-juu.**  
 eat-ADJR              be-1pl-PST  
*We would make the flour soup and the fried vegetables and eat.*  
 T0021:004



## 6 Switch Reference

Within the context of subject reference, Haspelmath (1995: 10) describes three types of converbs:

- those with implicit subjects, where the subject cannot be overt;
- those with explicit subjects, where the subject must be overt;
- those with free subjects, where the subject may be overt, but is not required to be.

In Darma we find that the subjects are free: They may be overt, but they are not required to be overt. We find an overt subject/agent in both types of converb construction.

We find the plain converb with an overt pronoun as in (14) below.

- (14) **ji k<sup>h</sup>wi tɔŋ-lən jɛr-hi.**  
 1sg dog see-CONV be\_afraid-1sg.PRS  
*After seeing the dog, I got scared.* T0042 Elicited: 338

We also find an overt subject/agent in the converb construction with the [-jɯ] marker, as in example (8) above. Looking at lines (ii) through (v), we find that the converb ‘start’ (e.g. light a fire) has an overt agent, ‘he’, and that the converb ‘catch’ also has an overt agent, ‘the police’, who catch the hero of the story and deliver him to the king. Referring back to (3) above, we find a verb stem with a converb morpheme and no overt subject/agent. The final verb in (3), ‘go’, is marked for first person plural present.

In example (15), we have another construction with a transitive verb stem marked with a converb morpheme, but here the transitive verb has no overt agent. The verb of the matrix clause is marked as first person plural, but an overt nominal agent is not required in either clause.

- (15) **hã re ru kur-lən-jɯ kuŋ ru**  
 then field in take\_away-CONV-PST grave in  
**lob ni-n-jɯ.**  
 bury be-1pl-PST  
*Then after we take the body to the field we bury it in a grave.* T0018:005

That a subject is not required in the matrix clause or in the subordinate converbal clause raises the question: How do we know who is doing what in a narrative? How do I know who is bringing the body to the field and

burying it in example (15)? One clue is on the finite verb. In Darma, 1pl, 2sg/pl agreement is encoded on the verb. For 3sg and 3pl there is no overt agreement marker, but the tense marker is distinct. This is shown in Table 1.

|            |                       |            |                 |
|------------|-----------------------|------------|-----------------|
| 1sg        | ja-Ø-jo <sup>12</sup> | <b>1pl</b> | <b>ja-n-cuu</b> |
| <b>2sg</b> | <b>ja-n-cuu</b>       | <b>2pl</b> | <b>ja-n-cuu</b> |
| 3sg        | ja-Ø-suu              | 3pl        | ja-Ø-suu        |

Table 1: A paradigm for ‘eat’

The verb takes the same marker for 1pl and 2sg/pl, while the verb form for 1sg is distinct in Darma, and 3sg and 3pl verb forms are identical.

According to Haspelmath (1995), if the converbal construction has no overt subject, the subject of the matrix clause is the subject of both clauses, which we find in example (2).

In (2) there is not an overt subject in either the matrix or subordinate clause. This does not pose a problem, however, because the matrix verb has a 1pl agreement marker after the stem. Thus, we know that the subject of both clauses is ‘we’.

This type of encoding works quite well if the subject of both clauses is the same. But what about sentences with multiple converbal constructions where each clause has a different subject, such as example (8)? The utterance starts out with the hero having a mischievous thought, then he sets the mill on fire, then the police catch him, and he is taken to the king. Each of the converb markers is followed by the [-ju] marker, which I have labeled past tense. Upon closer investigation, this marker appears to be what Haspelmath (1995) and Longacre (1985) call a ‘switch reference’ marker.

Longacre includes switch reference as part of his definition of a ‘clause chain’. Haspelmath does not go this far, but certainly includes a discussion of ‘switch reference’ in his discussion of converbal constructions. A ‘switch reference’ marker alerts the interlocutor to a change of subject for the next clause. So, in example (8) we start off talking about the hero of the story, and then talk about the police catching him. To indicate that the agent or subject is changing, the converb bears the ‘switch reference’ marker [-ju]. It should not be alarming that the first clause has our hero as a subject and the converb has the ‘switch reference’ marker on it even though there is not switch. In natural discourse there are false starts and stops, and this could be an example of an intended switch that doesn’t take place. The speaker states that the hero has a mischievous idea, and before switching to the

---

<sup>12</sup> The verb is in the following format: STEM-NUMBER-PERSON.

topic of the police, as planned, he decides to state that the hero set the mill on fire, so he says ‘he lights the mill on fire’ before shifting to a discussion of the police.

Another example of this phenomenon is given in (16). Based on context, we know that the utterance starts off in the first person plural. Following the converb [de-lən-ʃaŋ], the subject switches to the third person singular.

- (16) **hã**            **bakca**            **de-lən-ʃaŋ**            **ido sũ,**  
 then            wedding            go-CONV-PST            after\_that
- ta**    **ʃahi**            **le-ni,**            **niŋ-gu**    **ʃahi**    **do**  
 one    relative    be-3sg.PRS    1pl-POSS    relative    here
- ru**    **niŋ-gu**    **ts<sup>h</sup>etəŋ**            **ga-da.**  
 LOC    1pl-POSS    alter            do-3sg.PRS

*Then after going to the wedding, after that there is one relative,  
 our relative from here does the alter.*

T0013:014

Reviewing the data, I find that the pattern for switch reference is most robust in narrative discourse. Elicited utterances sometimes contradict the hypothesis that [-ʃu] is a switch reference marker. While I find examples of switch reference in the data examined, many examples that contain the converb + [-ʃu] construction have no switch of referent. This is an issue that I will address in the next section.

## 7 The ‘Switch Reference’ Marker [-ʃu] that Defies the Generalization

Informal discourse has the complicating factor of being spontaneous speech. We often start to say something and then decide to say something else mid-sentence. This may be a factor in cases where [-ʃu] does not actually mark a switch in natural discourse texts. While I was sensitive to the notion of pauses and other cues in the recorded texts as I worked with my native-speaker consultant, I was not considering an analysis of the [-ʃu] as a switch reference marker in these contexts. In fact, reflecting on interviews with my consultant, he was not cognizant of the role of the [-ʃu] marker in a way that allowed him to explicate its use. He provided several examples during elicitation sessions that I interpreted as contradictory; these examples have confounded my understanding of [-ʃu] in converbal constructions. My consultant is a native speaker of Darma and speaks the language



and that the marker on ‘catch’ in (8) cannot reasonably be termed a switch reference marker if the switch refers to the subject.

Another sentence to consider in this discussion is (18):

- (18) **niŋ jaʔi ga-lən ʔui jani.**  
 1pl food make-CONV 3pl eat-3pl.PRS  
*Upon our making the food they eat.* T0042 Elicited: 524

In (18), we find a 1pl subject in the converbal clause and a 3pl subject in the matrix clause. This is a construction where we would expect to find a switch reference marker on the converb, but we do not. We could hypothesize that if both subjects/agents are overt, then there is no need for a switch reference marker. This hypothesis can be quickly rejected if we recall that in (8) above we found overt subjects with the switch reference marker [-ʔu]. In (8), both 3sg ERG and ‘police’ ERG are overt, and yet we find the [-ʔu] marker in both clauses.

## 8 The Properties of Switch-Reference Revisited

While I have data to support the notion that [-ʔu] is a switch-reference marker, I have a lot of data that demonstrates that switch reference cannot be the only role of this morpheme. The fact that it does not always indicate a switch is problematic, but any account of the [-ʔu] marker must also account for the apparent switch in some constructions. In an effort to find a solution to this puzzle, let us look to Central Pomo (CP) and the analysis provided by Mithun (1993).

CP has three pairs of markers that appear to mark switch-reference. The role as a switch-reference marker is most evident in elicited constructions. In natural speech, however, the markers don’t always follow the same pattern (Mithun 1993). Upon further investigation, Mithun concludes that the markers actually encode how many events are taking place under the matrix clause; switch-reference as classically defined is not the primary function of these morphemes.

She demonstrates with the three pairs of markers in CP that each switch reference morpheme that encodes the same subject in the matrix and dependent clauses is used in constructions where the action of the dependent clause and the action of the matrix clause are together interpreted as being one event. The morpheme that marks switch-reference is used when the event in the dependent clause and the event in the matrix clause are two separate events. The primary function of the switch reference marker is not to encode a switch of referent in terms of subject/agent, but rather the marker is marking ‘same versus different eventhood’ (Mithun 1993: 134).

The examples<sup>13</sup> in (19) (Mithun 1993: 126) exemplify her argument:

- (19) **ʔá mk<sup>h</sup>e k<sup>h</sup>ʒéʔel dóčhi**  
 ʔá mk<sup>h</sup>e k<sup>h</sup>ʒé=ʔel dó-č-hi  
 1.AGT 2.AGT bridge=the make-SML-SAME
- míli ma ʔdímk<sup>h</sup>e**  
 mí = li ma ʔdí-m=ʔk<sup>h</sup>e  
 that=with 2.PAT take.PL-across=FUT

*I will build the bridge for you and on that you'll take them (across)  
 = I will build you a bridge to take them across on.*

- (20) **m̄to mul smá m̄tíčkayahla**  
 m̄to mul smá m̄tí-č-ka-ya=hla  
 2.PAT that sleep lie-INCH-CAUS-DEFOC=DIFF
- m̄to q'óʔti madúmač'k<sup>h</sup>e t<sup>h</sup>in**  
 m̄to q'ó = ʔti madú-mač'=k<sup>h</sup>e t<sup>h</sup>i-n  
 2.PAT what=even awake-AFF-INCH=FUT not-IP

*If you were put to sleep, you'd never wake up at all.*

In example (20) the subject is the same in both clauses, yet we find a switch reference marker on the medial verb.

The switch reference markers in CP are not found in the same constructions as the potential switch reference marker in Darma. The markers termed switch-reference markers according to Mithun do not occur as part of a converb construction. In fact, Mithun states that the switch-reference markers themselves play a role in the linking of clauses in CP (1993: 119). However these morphemes are identified, they are found in dependent clauses<sup>14</sup> and can either mark a switch of reference or mark the event in the dependent clause as distinct from the event in the matrix clause. Also, Darma does not have an alternation between a 'same' marker and a 'different' marker as Mithun found in CP. Despite these differences, let us revisit the Darma data and reanalyze it following the pattern found in CP. In our

<sup>13</sup> Mithun uses the following abbreviations: AFF = emotional affect; AGT = agent; CAUS = causative; DEFOC = defocus; DIFF = different—distinct events; FUT = future; INCH = inchoative; IP = imperfective aspect; PAT = patient; PL = plural; SAME = same—simultaneous actions; SML = semelfactive.

<sup>14</sup> While Mithun does not provide tests for subordination, and claims that these constructions may not be semantically subordinate, she does state that the clauses marked with the switch reference marker are 'grammatically dependent insofar as speakers do not feel they constitute complete sentences in themselves' (124).

reanalysis we will consider the switch reference marker [-ʃu] in Darma to be a morpheme that relates events rather than subjects. We will consider the overt marker to index different events. Under this analysis, I claim that the overt ‘different events’ marker [-ʃu] alternates with a zero marker that indexes ‘same events’.

## 9 Switch-Reference in Darma Revisited

Let us reconsider the example from (8) above that was problematic under each analysis considered to this point. In this example, the first two clauses have the same overt referent (3sg referring to the hero of the story), and yet we find the overt switch reference marker on the converb.

If we reanalyze the switch reference marker according to Mithun’s analysis, we can reexamine the utterance to consist of four separate events. First, the hero has a mischievous idea, the second event is that he lights the fire, the third event is the police capture the hero, and the final event is that the police deliver the hero to the king.

This analysis will allow us to explain why we find the switch reference marker in constructions where the subject remains the same in the matrix and subordinate clauses as we saw above in (15), where the subject/agent in both clauses is the first person plural, but the switch reference marker is overt. We now understand that the [-ʃu] is there to indicate that the taking away and the burying are two separate events.

When looking at an example like (11) using the new analysis of the switch reference marker, a new question arises: Why is there no switch reference marker on the converb ‘make’?

There is a cultural explanation for this. In the Rang culture, cooking and eating are one event. If one is invited to a traditional Rang house, the eating is done around the hearth. The women of the household cook the food and serve it to the guests seated around the periphery of the kitchen. With this in mind, we can posit a semantic interpretation of the events of cooking and eating as one event. This semantic interpretation is coded on the converb as a ‘same’ switch reference marker, which is morphologically unmarked.

Earlier I mentioned that eliciting these constructions in interview sessions was challenging. Mithun reports that elicitation sessions did not reveal the pattern found in discourse. In interview sessions with my primary consultant, BSS, I was told that the converb and the converb + switch reference constructions were the same. Examining the data, I find that BSS did not use these constructions interchangeably. That he would report that these constructions are used in the same way when asked about them in the abstract is not surprising. Without context, it is difficult to articulate how the

plain converb is different from the converb + switch reference marker construction.

The pattern of switch reference appears to be consistent in the Darma texts analyzed to date, but there are some constructions that do not fit neatly into the analysis. For example, (8) and (17) don't fit the analysis as I have presented it here.

These two examples both have the verb 'catch' as a converb, (8) with the switch reference marker and (17) without. They both also have the fully inflected verb 'take away'. The question is why does (8) have the switch reference, and (17) does not? It may be because the tense and aspect features of the matrix verb are different.<sup>15</sup> The CP data presented in Mithun (1993) have three sets of markers one pair for realis events that coincide, one pair for irrealis and the third pair for realis events that are consecutive. It seems logical that Darma would also distinguish between these different aspectual cases by using the switch reference differently in differing aspectual constructions. Another factor to consider is that the matrix verb construction in (8) is a compound verb (the stem 'deliver' with the fully inflected 'take away'), while the matrix verb in (17) is not. This type of construction with compound verbs and alternations with tense/aspect, which here contradicts the switch reference analysis, must be further examined.

## 10 Concluding Remarks

The converbal construction in Darma adheres to the definition from Haspelmath (1995) outlined here. The converb morpheme is a verb suffix; and the resulting converbal clause is subordinate to the matrix clause. The hypothesis of a switch reference marker poses two problems: First, it appears to be a time marker, which flouts Haspelmath's definition of a converb; second, the marker itself does not function primarily as a switch reference marker as the term is traditionally defined by Longacre, Haspelmath, and others.

Examining the data available in Darma, I suggest that Haspelmath's (1995) notion of a converb needs to be revised. I am not the first to suggest this. Genetti (2005) suggests that Haspelmath's definition should be revised; specifically, she suggests that his requirement of subordination be relaxed. Additionally, cross-linguistic evidence supports this suggestion that the definition of the term 'converb' should be revised. While Longacre terms the constructions presented here 'clause chains', he does not require subordination outright. Rather he suggests that the multiple verbs in these constructions are not of the same rank. The medial verbs are not structurally

---

<sup>15</sup> Thank you to Anthony C. Woodbury for suggesting that tense and aspect may play a role in the use of the switch reference morpheme.

the same as the final verb, which is fully inflected. In the Darma construction only the final verb is fully inflected with tense, aspect, and agreement; the preceding verbs are maximally the stem with the nonfinite morpheme and the past tense marker, which serves as a switch reference marker. Identifying a switch reference marker that does not index a switch of subject is also justified cross-linguistically. Darma and Central Pomo utilize the marker in a different sense, marking a switch of event rather than subject.

I propose that we maintain the use of ‘converb’ and ‘switch reference’ to discuss these morphemes as Haspelmath and Koenig (1995) and others have established in the literature. We should expand the definition of these terms to account for data found cross-linguistically. To date, I have found one other language, Caodeng rGyalorong (Sun 2003), that uses a past tense in converb constructions. The construction is not the same as Darma; the past form of the verb stem is reduplicated to form the converb. But it is clear that excluding tense from the converb is not the pattern found cross-linguistically. I must continue a search of the literature to see if I can locate other languages where past tense is a part of the converbal construction.

While I propose a revision of Haspelmath’s definition, I do support continuing the use of the term ‘converb’ to categorize these constructions. Utilizing the same terminology will allow us to assert that, while the details of how the morpheme functions in a given language differ, the role it plays in general follows a pattern found cross-linguistically. The converb marks a nonfinite verb within a larger structure. A switch reference marker indexes a shift in subjecthood/eventhood within a larger structure.

Finally, I must continue to test this analysis in Darma. I must finish analyzing the remaining texts under the analysis presented here and see how robust the pattern is.

## References

- Census. 2005. 2001 Census. Office of the Registrar General, India.
- Genetti, C. 2005. The Participial Construction of Dolak<sup>h</sup>a Newar. *Studies in Language* 29:35-87.
- Gordon, R.G. 2005. *Ethnologue: Languages of the World*, 15<sup>th</sup> edition. SIL International.
- Haspelmath, M. 1995. The Converb as a Cross-linguistically Valid Category. *Converbs in Cross-Linguistic Perspective*, eds. M. Haspelmath and E. Koenig, 1-55. Berlin: Mouton de Gruyter.
- Haspelmath, M. 2002. *Understanding Morphology*. London: Arnold.
- Haspelmath, M. and E. Koenig (Eds.). 1995. *Converbs in Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter.

- Krishan, S. 2001. Darma, Chaudangsi, and Raji. *New Research on Zhangzhung and Related Himalayan Languages*, ed. R. LaPolla, 343-501. Osaka: National Museum of Ethnology.
- Longacre, R.E. 1985. Sentences as Combinations of Clauses. *Language Typology and Syntactic Description*, ed. T. Shopen, 235-286. Cambridge: Cambridge University Press.
- Masica, C.P. 2005. *Defining a Linguistic Area: South Asia*. New Delhi: Chronicle Books.
- Mithun, M. 1993. 'Switch-Reference': Clause Combining in Central Pomo. *International Journal of American Linguistics* 59:119-136.
- Saxena, A. 2004. On Discourse Functions of the Finite Verb in Kinnauri Narratives. *Himalayan Languages: Past and Present*, ed. A. Saxena, 213-236. Berlin: Mouton de Gruyter.
- Snell, R. 1992. *Teach Yourself Hindi*. Chicago: NTC/Contemporary Publishing.
- Sun, J.T.-S. 2003. Caodeng rGyalrong. *The Sino-Tibetan Languages*, ed. G. Thurgood and R. J. LaPolla, 490-502. New York: Routledge.
- Trivedi, G.M. 1991. *Descriptive Grammar of Byansi: A Bhotiya Language*. Calcutta: Anthropological Survey of India.