Morphotactics in an Information-Based Model of Realisational Morphology

Olivier Bonami
Université Paris-Sorbonne
Institut universitaire de France
LLF (UMR 7110)

Berthold Crysmann
CNRS
LLF (UMR 7110)

Proceedings of the 20th International Conference on
Head-Driven Phrase Structure Grammar

Freie Universität Berlin

Stefan Müller (Editor)

2013

CSLI Publications

pages 27–47

http://csli-publications.stanford.edu/HPSG/2013

Bonami, Olivier, & Crysmann, Berthold. 2013. Morphotactics in an Information-Based Model of Realisational Morphology. In Müller, Stefan (Ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, 27–47. Stanford, CA: CSLI Publications.

Abstract

In most recent work, Crysmann and Bonami (2012) suggest to reconcile the insights of inferential-realisational morphology (Anderson, 1992; Stump, 2001; Brown and Hippisley, 2012) with the full typology of variable morphotactics: situations where the expression of analogous feature sets can appear in various positions in the string. The authors proposed to account for these facts by importing, into HPSG, a variant of Paradigm Function Morphology (Stump, 2001) where realisation rules are doubly indexed for linear position and paradigmatic opposition. In this paper we first introduce more empirical challenges for theories of morphotactics that neither PFM nor the reformist approach of Crysmann and Bonami (2012) can accommodate. We then argue for a reappraisal of methods for morph introduction, and propose a new approach that replaces stipulation of classes of paradigmatic opposition with a general distinction between expression and conditioning (Carstairs, 1987; Noyer, 1992) which greatly expands the scope of Pāṇini's Principle.

1 Variable morph ordering

1.1 Types of non-canonical morphotactics

In the inflection of a particular lexical category in a given language, morphs are most canonically organised in a sequence of POSITION CLASSES: morphs expressing different values for the same feature cluster in a single linear position, strictly ordered with respect to positions serving for the realisation of other features.[1]. Of course, deviations from this canonical ideal are very common, and come in many varieties; most well-known are FUSED EXPONENCE (a single position realises more than one feature), EXTENDED EXPONENCE (the same feature is realised simultaneously in multiple positions), and ZERO EXPONENCE (some feature is not expressed at all); these famously motivate the Word and Paradigm family of approaches to inflection (Matthews, 1972).

A family a deviations of particular interest is that of VARIABLE MORPH ORDERING. This again comes in multiple varieties. In POSITIONAL DISAMBIGUATION, the same morph expresses related but distinct morphosyntactic property sets in different positions. A nice example is that of subject and object markers in Swahili (Stump, 1993),

[†]We would like to thank the audience of HPSG 2013, in particular Doug Arnold, Rob Malouf, Jesse Tseng, and Frank van Eynde. We are also greatly indebted to Greg Stump for his highly stimulating reactions, suggestions and constructive criticism he shared with us on various occasions. Previous version of the present paper have also been presented at the First American International Morphology Meeting 2012 at Amherst and at a seminar at the University of Essex. We therefore would like to extend our thanks to the respective audiences of these two venues, in particular to Farrell Ackerman, Mark Aronoff, Louisa Sadler, and Andrew Spencer. Of course, the usual disclaimers apply. This work is related to work package *Morph1* of Labex EFL (funded by the ANR/CGI). Participation at HPSG 2013 was supported by the project TranSem, funded by the CNRS programme PEPS HuMaIn.

Authors' names are listed in alphabetical order, sorted by last name for a change.

[1]Whether the relative order of these positions should be assumed to be canonically correlated with the identity of the features (Bybee, 1985; Rice, 2000; Aronoff and Xu, 2010) is debatable. However we assume without discussion, with e.g. Stump (1993, 2001); Nordlinger (2010) that the relative order is often arbitrary, and we focus on the treatment of the arbitrary cases.

as illustrated in Table 1: these markers are homophonous for most nominal classes, but do not appear in the same position within the verb.

PER	GEN	SUBJECT SG	SUBJECT PL	OBJECT SG	OBJECT PL
1		ni	tu	ni	tu
2		u	m	ku	wa
3	M/WA	a	wa	m	wa
	M/MI	u	i	u	i
	KI/VI	ki	vi	ki	vi
	JI/MA	li	ya	li	ya
	N/N	i	zi	i	zi
	U	u	—	u	—
	U/N	u	zi	u	zi
	KU	ku	—	ku	—

Table 1: Subject and object prefixes in Swahili

In MISALIGNED EXPONENCE, morphs that are in paradigmatic opposition appear in different linear positions. The Laz subject markers in Table 2 exemplify (Lacroix, 2009): with intransitive verbs, subject agreement is marked suffixally by default, prefixally in the first person, and both prefixally and suffixally in the 1PL.

LAL 'bark'		
	SG	PL
1	b-lalum	b-lalum-t
2	lalum	lalum-t
3	lalum-s	lalum-an

Table 2: Subject marking on simple intransitive verbs in Laz

In CONDITIONED REORDERING, one and the same morph expressing the same property set appears in different linear positions depending on some (phonological, morphosyntactic, or semantic) condition.[2] Mari nominal declension offers a relevant example (Luutonen, 1997), as shown in Table 3: in the accusative, the possessor marker precedes the case marker, while in the lative, it is the other way round. In FREE REORDERING, the expression of some combination of morphosyntactic properties relies on two morphs whose relative order is not constrained by the grammar. This is also found in Mari declension, in the dative.

From a theoretical point of view, the Mari data are highly informative, since they actually provide the missing typological link between free ordering, as observed for Chintang (Bickel et al., 2007) and conditioned reordering, as manifest in Laz or Fula (Stump, 1993): systems that feature essentially free permutation, but are constrained for some cells, lend themselves quite naturally to an analysis in frameworks that build

[2]Reversible and ambifixal position classes (Stump, 1993) are two subcases of conditioned reordering.

	NOPOSS	1PL.POSS	
		POSS ≺ CASE	CASE ≺ POSS
NOM	пӧрт	пӧрт-на	
ACC	пӧрт-ым	пӧрт-на-м	*
DAT	пӧрт-лан	пӧрт-на-лан	пӧрт-лан-на
LAT	пӧрт-еш	*	пӧрт-еш-на

Table 3: Partial paradigm of Mari possessed nouns (Riese et al., 2010)

on the accumulation of partial descriptions.

1.2 Approaches to variable morphotactics

Within the Word and Paradigm tradition, the most prevalent view of morphotactics rests on three crucial assumptions (Anderson, 1992; Stump, 2001): (i) morphological composition is stem-centric: it starts from the lexeme's basic stem which it modifies incrementally by sequential application of morpholexical rules; (ii) morpholexical rules operate on morphologically unstructured ('a-morphous') phonological representations; (iii) morpholexical rules are organised into blocks of mutual exclusivity.

This set of assumptions gives rise to a view of morphotactics where exponents in paradigmatic opposition are expected to linearise in onion-like fashion, as outlined in Fig. 1. Deviations from this expectation have been recognised early on, and dealt with using different analytic devices over the years: metarules (Stump, 1993), rules of referral (Stump, 2001), or conditional operators of composition and linearisation (Stump, 2012a,b). Still, all these proposals share the view that the kind of morphotactic structure illustrated by Fig. 1 is the least marked.

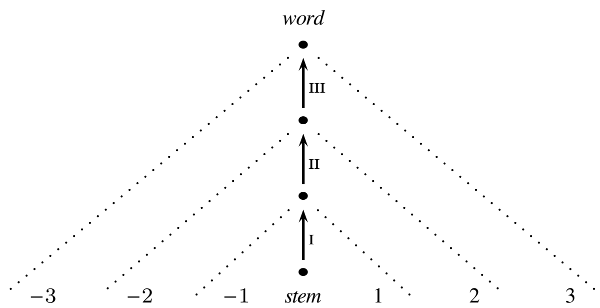

Figure 1: The interaction of rule blocks and morphotactics

Crysmann and Bonami (2012) challenge this assumption, and argue that the types of variable morphotactics found in the languages of the world do not warrant taking any type of variable morphotactics as less marked: what is less marked is to not have any variation in order, but there is no preference for variations that occur in onion-like fashion around the stem. Crucial to their argumentation is the pattern commonly found with Romance pronominal affixes, here illustrated with Italian (Monachesi, 1999), where sequences of affixes occur in the same order on either side of the stem (see also Luís and Spencer, 2005).

(1) a. me-lo-dai
DAT.1SG-ACC.3SG.M-give.PRS.2SG
'You give it to me.'

b. dá-me-lo!
give.IMP.2SG-DAT.1SG-ACC.3SG
'Give it to me!'

Crysmann and Bonami (2012) propose a reformist modification of standard assumptions amounting to dropping (i) above: instead of licensing inflected words starting from the stem, they start from the left edge of the word and delay the introduction of the stem. Morpholexical rules carry an explicit position class index, and order variability is dealt with by underspecification of position class. Crucially, this analytic setup enables Crysmann and Bonami to deal with a wider typology of variable morphotactics while keeping two central analytic assumptions: the a-morphous hypothesis and the organisation of rules into blocks.

2 New challenges

2.1 Rule blocks and position classes

While Crysmann and Bonami (2012) arguably provides for a more refined theory of variable morphotactics than its predecessors, the proposed formal analysis is a hybrid, which ends up having unsatisfactory design properties. Particularly inelegant is the double indexing of morpholexical rules for rule blocks (encoding paradigmatic opposition) and position class (encoding syntagmatic order). While some indexing scheme for positions is indisputably necessary, the necessity of block indices is far less clear in a system where these indices are dissociated from linear order.

When stripped of their function of deriving linear order, what rule blocks appear to do is just ascertain morphological wellformedness: in inferential-realisational models of morphology and constraint-based grammar alike, a bare stem, being underspecified, may denote any cell of the paradigm, the only problem being that such a stem more often than not fails to constitute a legitimate morphologically well-formed word. Thus, one of the two remaining functions of rule blocks is to ensure that any inflectional feature that has some expression must be expressed and that cases of zero exponence are limited to the cells in the paradigm for which the system provides no exponent. The other remaining function for rule blocks is to limit the scope of Pāṇinian competition, in order to permit instances of extended exponence, i.e., multiple expression of same or overlapping morphosyntactic properties (see section 3.2). But if rule blocks are divorced from the expression of constraints on order, one ends up with a completely unconstrained way of exempting morphs from Pāṇinian competition, severely undermining the status of Pāṇinian competition as an organisational principle of morphological systems.

Another problematic issue with Crysmann and Bonami (2012) — actually a defect inherited from the PFM model — concerns their treatment of the identity function default (*ifd*), i.e., the morpholexical rule to account for zero exponence: ideally, there should only be one such default realisation rule that captures every morphosyntactic property that does not have an independent overt realisation. However, owing to the logic of rule blocks, Crysmann and Bonami (2012) need to postulate not only multiple instances of the same default rule, but also need to ensure that such an instance exists for every rule block.

More generally, within the context of information-based syntax and semantics, the idea of stipulating a system of ordered or unordered rule blocks merely for the purposes of ensuring morphological wellformedness should come as a bit of an embarrassment, even more so, if wellformedness can be simply captured by a straightforward principle: every property that can be expressed, needs to be expressed. In this paper, we shall develop a model of realisational morphology within HPSG that replaces stipulated static blocks of paradigmatic opposition with a general principle that manages the expression of morphosyntactic resources. We will show that this approach is not only preferable on a conceptual level, but also supported by an increase in analytical elegance and empirical coverage.

2.2 Challenging a-morphousness

Wackernagel affixes are affixes that are constrained to be the second realised morph in the word (Nevis and Joseph, 1992). A clear example is provided by Sorani Kurdish (Samvelian, 2007) and illustrated in Table 4. In past transitive verbs, if the verb is VP initial, the set of markers realising subject agreement are realised immediately after the first other morph, irrespective of whether that morph is the basic stem, a negative prefix, or an aspectual prefix.[3]

1	2	3	4	
		nard=**jân**	im	‘they sent me’
na=**jân**		nard	im	‘they did not send me’
	da=**jân**	nard	im	‘they were sending me’
na=**jân**	da	nard	im	‘they were not sending me’

Table 4: Sorani Kurdish past person markers

Wackernagel affixes pose a serious challenge to the a-morphousness assumption. In order to know where to linearise the affix, one needs to keep track of the position of the first overtly realised morph in the word. Both in conventional stem-centric approaches and in Crysmann and Bonami (2012)'s left-to-right approach, this information is inaccessible: the morpholexical rule introducing **jân** can only access phonological properties of its input, not morphological properties; thus irrespective of the order in which rules apply, there is no way of checking what the morphological structure of

[3]See section 4 for a more detailed description.

the sequence on the right of the affix is.[4] Stump (2012b) circumvents this problem by redefining realisation rules so that they construct two phonological strings in parallel: in addition to the full phonological representation, realisation rules recursively define the *pivot*, the substring of the whole phonology at whose edge Wackernagel affixes are to be realised. We would argue that this amounts to abandoning the spirit, if not the letter, of the amorphous hypothesis: using morphologically segmented phonological strings or recording separately the location of morph boundaries are just two equivalent ways of remembering where those boundaries are. In the remainder of this paper we suggest a more direct approach to this phenomenology.

3 Analysis

3.1 Information-based realisational morphology

Inferential-realisational models of morphology typically draw a distinction between morpholexical rules (or realisation rules), which provide recipes for the introduction of exponents, and a system of paradigm functions that concert the way in which these recipes are applied to yield a well-formed word. In a-morphous approaches, such as Anderson (1992); Stump (2001); Crysmann and Bonami (2012), morpholexical rules are formulated as (potentially recursive) unary rules. Paradigm functions then guarantee that exactly the right number of rules are invoked, in the right order. Choice between competing rules is currently understood as being governed by Pāṇinian competition. This division of labour between morpholexical rules and paradigm function has proven quite successful, since it permits reuse of resources, as needed, e.g., for the treatment of positional disambiguation (cf. table 1; Stump, 1993; Crysmann and Bonami, 2012).

While keeping this general division, we shall revise the formal nature of morpholexical rules: instead of rule cascades successively transforming a basic stem into a complete word, rules will be considered instead as pairings between morphosyntactic properties and lists of exponents. Building on ideas proposed in Crysmann (2002), we postulate a flat structure of segmentable morphs (not morphemes) which are indexed for position. In essence, we are moving structure away from the derivation history into morphological representations. This move actually provides for a more restrictive model, since it systematically disallows reference to the derivation history.

Morpholexical rules are represented by feature structures organised in a type hierarchy, providing a pairing of a list of MORPHS with the morphosyntactic features they express (M(ORPHOLOGY) U(NDER) D(ISCUSSION)). In order to capture allomorphic conditioning, morpholexical rules may impose constraints on morphosyntactic properties they do not strictly realise, e.g. the negative allomorph of the Swahili past marker in

[4]This is not literally true of Crysmann and Bonami (2012)'s approach, because of a technical defect in the formulation of realisation rules: the authors code recursion of realisation rules by the HPSG-standard use of a DTR feature. Thus in fact the whole derivation history is accessible to later rule application. This is clearly a poor design choice that does not correspond to the intended a-morphous interpretation of rules.

(2). The morphs thus introduced consist of a phonological description (PH) together with a position class index (PC). The formal encoding of position class indices is discussed in the appendix.

(2) $$\begin{bmatrix} \text{MUD} & \boxed{1}\{\textit{past}\} \\ \text{MORSYN} & \boxed{1} \cup \{\textit{neg}, ...\} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \langle \text{ku} \rangle \\ \text{PC} & 3 \end{bmatrix} \right\rangle \end{bmatrix}$$

Since morpholexical rules "know" what features they express (MUD), we can define morphological completeness and coherence in terms of resource consumption: as stated in (3), the morphosyntactic features expressed by morphological rules must match up to produce the morphosyntactic property set of the word. As for exponence, we compute the MORPHS list of a word by shuffling all the morphs contributed by the morpholexical rules in the order of their position class indices.[5]

(3) $$\textit{word} \rightarrow \begin{bmatrix} \text{MORPHS} & \boxed{e_1} \bigcirc \cdots \bigcirc \boxed{e_n} \\ \text{MORSYN} & \boxed{0} \ (\boxed{m_1} \uplus \cdots \uplus \boxed{m_n}) \\ \text{RULES} & \left\langle \begin{bmatrix} \text{MORPHS} & \boxed{e_1} \\ \text{MUD} & \boxed{m_1} \\ \text{MORSYN} & \boxed{0} \end{bmatrix}, \ldots, \begin{bmatrix} \text{MORPHS} & \boxed{e_n} \\ \text{MUD} & \boxed{m_n} \\ \text{MORSYN} & \boxed{0} \end{bmatrix} \right\rangle \end{bmatrix}$$

Compared to Crysmann and Bonami (2012), position class information is considered a property of the morphs here, rather than a property of rules. Note further that there are no rule block indices (or POI), ensuring morphological completeness and coherence entirely in terms of the principle in (3). As a direct consequence, we extend the scope of Pāṇinian competition to all maximally specific types:

(4) a. For any leaf type t_1[MUD μ_1,MORSYN σ], t_2[MUD μ_2,MORSYN $\sigma \wedge \tau$] is a morphological competitor, iff $\mu_1 \subseteq \mu_2$.

b. For any leaf type t_1 with competitor t_2, expand t_1's MORSYN σ with the negation of t_2's MORSYN $\sigma \wedge \tau$: $\sigma \wedge \neg(\sigma \wedge \tau) \equiv \sigma \wedge \neg\tau$.

Essentially, we formulate Pāṇini's Principle solely in terms of the information being expressed: morpholexical rules that express more properties (MUD) compete with those that express less, and those that have more specific conditioning (MORSYN) compete with those that are less strictly conditioned. Our version of Pāṇini's Principle has the further benefit that we only need a single instance of the identity function default (*ifd*), the morpholexical rule that deals with zero exponence:

[5] Morph lists are shuffled rather than simply concatenated because we want to allow a single rule to introduce two (or more) possibly discontinous morphs simultaneously: in such a situation a separate rule may introduce a morph in an intermediate position. "$\uplus$" denotes disjoint union: $X \uplus Y = X \cup Y$ if $X \cap Y = \emptyset$ and is undefined otherwise. Note that although we take RULES to be a list rather than a set, the relative order of elements in RULES currently plays no role in our analyses.

(5) $\begin{bmatrix} \text{MORPHS} & \langle\,\rangle \\ \text{MUD} & \boxed{1}\left\{[\;]\right\} \\ \text{MORSYN} & \boxed{1} \cup \textit{set} \end{bmatrix}$

Since the *ifd* specifies one completely underspecified MUD value, it is in competition with every other morpholexical rule, having its MORSYN value restricted to exactly those morphosyntactic features that do not have any independent expression, which is clearly a desirable result.

In the following two subsections, we show that this approach provides for a more general and less stipulative approach to competition and variable morphotactics.

3.2 Swahili negative marking

The first set of data we are going to investigate in detail pertains to Pāṇinian competition between different position classes and the treatment of extended exponence.

In Swahili, sentential negation is regularly marked by means of the prefix *ha* in slot 1 of the verb (cf. (6a)). However, if the verb is inflected for relative agreement, negation is expressed instead by the marker *si* in slot 3. Since *si* in (6b) is the only overt exponent of negative marking, we must conclude that negative relative *si* expresses negation, preempting the use of the regular negative marker *ha* (6c).

(6) a. **ha**- wa- ta- taka
NEG 3PL FUT want
'they will not want'

b. watu wa- **si-** o- soma
people 3PL NEG.REL REL.PL read
'people who do not read'

c. * watu **ha-** wa- ***(si-)** o- soma
people NEG 3PL NEG.REL REL.PL read

In PFM, where Pāṇini's principle is limited to individual rule blocks, and rule blocks are tied to linear position, there is no way to capture this directly. Under our purely information-based approach, preemption of *ha* by *si* follows directly given the proper subsumption of MORSYN specifications.

(7) a. $\begin{bmatrix} \text{MUD} & \{\textit{neg}\} \\ \text{MORSYN} & \textit{set} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \text{<ha>} \\ \text{PC} & 1 \end{bmatrix} \right\rangle \end{bmatrix}$

b. $$\begin{bmatrix} \text{MUD} & \{\mathit{neg}\} \\ \text{MORSYN} & \{\mathit{rel}\} \cup \mathit{set} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \text{<si>} \\ \text{PC} & 3 \end{bmatrix} \right\rangle \end{bmatrix}$$

Similarly, we can also derive competition between regular markers and portmanteaux without any stipulation in terms of rule block indices, contrary to Stump (1993) and Crysmann and Bonami (2012): as witnessed in (8), the 1SG negative portmanteau *si* simultaneously preempts the regular marker of negation *ha* and the regular marker of 1st singular subject agreement *ni*.

(8) a. (ha-) a- ta- ku- taka
NEG 3SG.SUBJ FUT 2SG.OBJ pay
'He will (not) pay you.'

b. (*ha-) ni- ta- ku- taka
NEG 1SG.SUBJ FUT 2SG.OBJ pay
'I will (*not) pay you.'

c. si- ta- ku- taka
NEG.1SG.SUBJ FUT 2SG.OBJ pay
'I will not pay you.'

Again, Pānini's principle directly accounts for preemption, based on the subset relation of MUD values. See the appendix on the simultaneous occupancy of two position classes.

(9) a. $$\begin{bmatrix} \text{MUD} & \left\{ \begin{bmatrix} \mathit{subj} & \\ \text{PER} & \mathit{1} \\ \text{NUM} & \mathit{sg} \end{bmatrix} \right\} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \text{<ni>} \\ \text{PC} & 2 \end{bmatrix} \right\rangle \end{bmatrix}$$

b. $$\begin{bmatrix} \text{MUD} & \left\{ \begin{matrix} \mathit{neg}, \\ \begin{bmatrix} \mathit{subj} & \\ \text{PER} & \mathit{1} \\ \text{NUM} & \mathit{sg} \end{bmatrix} \end{matrix} \right\} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \text{<si>} \\ \text{PC} & 1-2 \end{bmatrix} \right\rangle \end{bmatrix}$$

Having established how the extended domain of competition benefits the treatment of preemption across position classes, we shall now address how we integrate cases of extended exponence.

Consider the examples in (10): here, *ha* is clearly the only overt exponent of negation, so we can conclude that it actually expresses negation. In (11), we find extended exponence of negative marking, triggering the presence of *ha* together with a special negative past marker *ku*. However, since we have already established independently *ha* as the expression of negation, and furthermore, since negative past *ku* cannot independently signal negation, it follows that choice of the past marker is merely conditioned by negation.

(10) a. tu- ta- taka
1PL FUT want
'we will want'

b. **ha**- tu- ta- taka
NEG 1PL FUT want
'we will not want'

(11) a. tu- li- taka
1PL PST want
'we wanted'

b. *(**ha**-) tu- **ku**- taka
NEG 1PL PST.NEG want
'we did not want'

Drawing on our distinction between MUD and MORSYN, we can capture this situation straightforwardly:

(12) a.
$$\begin{bmatrix} \text{MUD} & \{\textit{past}\} \\ \text{MORSYN} & \textit{set} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \langle \text{li} \rangle \\ \text{PC} & 3 \end{bmatrix} \right\rangle \end{bmatrix}$$

b.
$$\begin{bmatrix} \text{MUD} & \{\textit{past}\} \\ \text{MORSYN} & \{\textit{neg}\} \cup \textit{set} \\ \text{MORPHS} & \left\langle \begin{bmatrix} \text{PH} & \langle \text{ku} \rangle \\ \text{PC} & 3 \end{bmatrix} \right\rangle \end{bmatrix}$$

Because *ku* is merely allomorphically conditioned on negation, it is not a competitor of *ha*, owing to disjoint MUD values. With respect to TAM marking, however, *ku* is a competitor of *li*, given identity of MUD and subsumption of MORSYN specification. Thus, based on a principled distinction between realising a property and being conditioned on some property (Carstairs, 1987), we can actually dispense with rule blocks and extend the scope of Pāṇini's Principle, without facing problems with extended exponence.

3.3 Mari declension (variable morphotactics)

To illustrate how the present account deals with reordering phenomena, let us turn back to the partial paradigm of Mari nouns illustrated in Table 3. This phenomenology is best described by stating that the relative order of case and possessor markers in Mari nominal declension in unconstrained by default; only specific case values call for one or the other order. This can easily be done within the current framework by underspecifying the position index of all possessor and some case affixes, only stating that it has to be higher than that of the stem, here 1. In the accusative (resp. lative), the possessor is forced to occur in position 2 (resp. 3) because the other position is already occupied by the case marker; in the dative, both orders are possible because neither affix is constrained to a specific slot. Arguably such a view is preferable to any view that arbitrarily chooses one relative ordering as basic and takes special measures to authorise reordering in particular instances (cf. e.g. the use of a conditional composition operator in Stump, 2012b).

(13) Variable position affixes

a. $$\begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{PC} & 1+n \\ \text{PH} & <\text{на}> \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} \textit{poss} & \\ \text{PER} & \mathit{1} \\ \text{NUM} & \mathit{pl} \end{bmatrix} \right\} \end{bmatrix}$$

b. $$\begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{PC} & 1+n \\ \text{PH} & <\text{лан}> \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} \text{CASE} & \mathit{dat} \end{bmatrix} \right\} \end{bmatrix}$$

(14) Fixed position affixes

a. $$\begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{PC} & 3 \\ \text{PH} & <\text{м}> \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} \text{CASE} & \mathit{acc} \end{bmatrix} \right\} \end{bmatrix}$$

b. $$\begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{PC} & 2 \\ \text{PH} & <\text{еш}> \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} \text{CASE} & \mathit{lat} \end{bmatrix} \right\} \end{bmatrix}$$

4 Edge-relative positioning of morphs

In this final section we illustrate how the current approach can deal with Wackernagel affixes. We first present in more detail the Sorani Kurdish data, and then outline an analysis that combines analytic innovations proposed by Crysmann and Bonami (2012) to deal with variable morphotactics with a new position indexing strategy that relies on reified morphs.

4.1 The Sorani Kurdish data

We first outline the system of person marking in Sorani Kurdish. Sorani Kurdish possesses two sets of person markers for verbs, which Bonami and Samvelian (2008) call respectively *verbal person endings* (VPEs) and *mobile person markers* (MPMs). The forms of these markers are indicated in table 5. The function of the markers is

	SG	PL
1	-im	-în
2	-î	-in
3	∅	-in

VPEs

	SG	PL
1	-im	-mân
2	-it	-mân
3	-î	-jân

MPMs

Table 5: The form of Sorani Kurdish person markers

TENSE	SUBJECT AGREEMENT	PRONOMINAL OBJECT
pres	VPE	—
	VPE	MPM
past	VPE	—
	MPM	VPE

Table 6: The distribution of Sorani Kurdish person markers

variable depending on the morphosyntactic context. In present tense, VPEs function as subject agreement markers (15a), whereas MPMs are object pronominal affixes (15b). In past tense, the situation is much more intricate. With strictly intransitive verbs, only VPEs are used, and they function as subject agreement markers (16). With transitive verbs the form-function mapping is reversed: MPMs now function as subject agreement (17a), and VPEs function as object pronominal affixes (17b). The situation is summarised in Table 6.

(15) a. Bâzirgân-akân asp-akân da-kir-in.
merchant-DEF.PL horse-DEF.PL IPFV-buy.PRS-3PL
'Narmin is buying the horses.'

b. Bâzirgân-akân da=jân=kir-in
merchant-DEF.PL IPFV=3PL=buy.PRS-3PL
'The merchants are buying them.'

(16) Bâzirgân-akân hât-in.
merchant-DEF.PL arrive.PST-3PL
'The merchants arrived.'

(17) a. (Ema) asp-akân=mân kirî.
1PL horse-DEF.PL=1PL buy.PST
'We bought the horses.'

b. (Ema) kirî=mân=in.
1PL buy.PST=1PL=3PL
'We bought them.'

Turning to morphotactics now, VPEs have a simple distribution: they occur in a fixed position to the right of the stem. MPMs exhibit a much more intricate pattern.

First, they behave as ENDOCLITICS (Harris, 2002). They are always realised on the word at the right edge of the first constituent of the VP (17a). In general, this means being realised as the last morph of that word. If however that word is a verb, then the MPM interacts with verb-internal morphotactics. By default it is the second morph in the word, as evidenced in Table 4. There are however some contexts where the MPM is realised instead in a fixed position to the right of VPEs: if the MPM is 3SG (18a) or if it is plural and cooccurs with a 1SG VPE (18b).[6]

(18) a. kirî-n-î.
buy.PST-3PL=3SG
'He bought them.'

b. kirî-m-tân.
buy.PST-1SG=2PL
'You (pl.) bought me.'

4.2 Analysis

Three ingredients are crucial to our account of this dataset.[7]

First, we account for the form-function reversal in the use of the two sets of person markers by appealing to an indirection between argument structure and MORSYN sets, as indicated in Figure 2. We assume that arguments indexed by inflectional morphology are coded in MORSYN using feature structures of two distinct types, *vpe* and *mpm*. The alignment between the representation of indexed arguments in ARG-ST (for the purposes of syntax) and MORSYN (for the purposes of morphology) varies depending on tense and transitivity. Specifically, intransitive verbs associate their sole direct argument with a *vpe* structure in MORSYN. For transitive verbs there are two strategies depending on tense: in the present the subject associates with a *vpe* structure and the object with an *mpm*, while the reverse situation is found in the past.

Second, to account for the default verb-internal positioning of MPMs, we need a way of explicitly making reference to the position of the first realised morph in a word. Crysmann and Bonami (2012) already recognised two indexing schemes for morphs. In *absolute indexing*, morphs are indexed in terms of absolute numbered position. Together with underspecification, this is sufficient to deal with most morph ordering situations, as illustrated above in the case of Mari. As Crysmann and Bonami (2012) show, the morphotactics of the Romance verb motivates the introduction of *stem-relative indexing*: under their analysis, in an Italian verb, the position of pronominal affixes is fixed, the position of the stem is underspecified, and the position of TAM and agreement markers is defined relative to the position of the stem. To this effect, a feature STM is introduced on morphs, that is shared by all morphs in the word:

(19) $$word \rightarrow \left[\text{MORPHS} \quad \left\langle \left[\text{STM } \boxed{s}\right], \left[\text{STM } \boxed{s}\right], \ldots, \left[\text{STM } \boxed{s}\right] \right\rangle \right]$$

[6]More fine points of Sorani morphotactics are discussed in Walther (2012).

[7]We focus on realisation of MPMs within the verb. Realisation at a distance can be dealt with using e.g. an edge feature mechanism, and is an issue orthogonal to our current concerns.

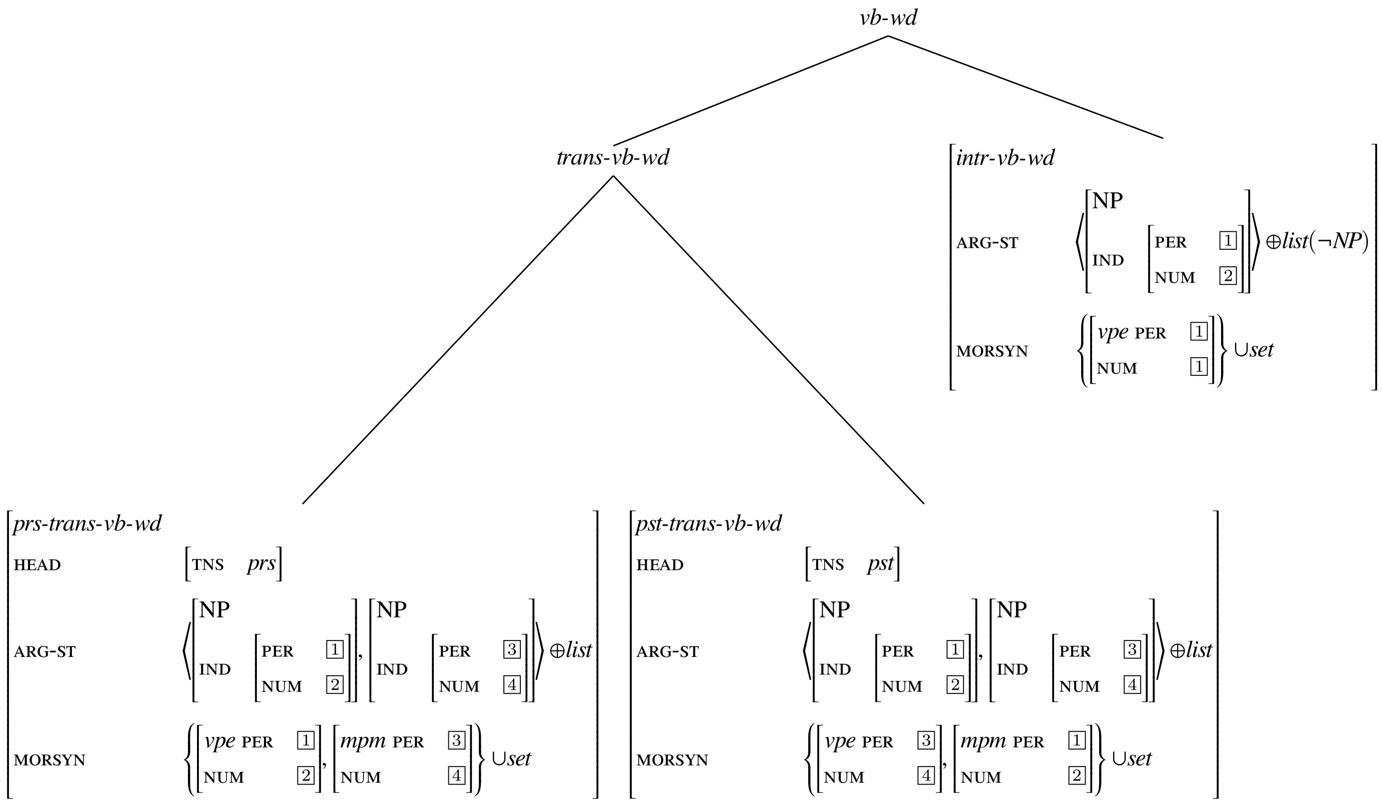

Figure 2: Types of alignment between grammatical functions and person markers

Morpholexical rules introducing a stem identify their own position index with that of the STM feature. Hence the position index of the stem is accessible to all morphs, and particular morpholexical rules may constrain the exponent they introduce to be realised at some fixed distance from the stem.

A variant of this analytic technique allows us to deal with Wackernagel affixes. We introduce a new feature 1ST that records the position index of the first realised morphs and makes that information available to other morphs (20). Thus a morpholexical rule for a second position affix such as the schematic rule in (21) will be able to indicate that the morph it introduces has to be adjacent to the first realised morph.

(20) $$word \rightarrow \begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{1ST} & \boxed{1} \\ \text{PC} & \boxed{1} \end{bmatrix}, \begin{bmatrix} \text{1ST}\ \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} \text{1ST}\ \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

(21) $$\begin{bmatrix} \text{MORPHS} & \left\langle \begin{bmatrix} \text{1ST} & \boxed{1} \\ \text{PC} & \boxed{1}+1 \end{bmatrix} \right\rangle \\ \text{MUD} & \left\{ \begin{bmatrix} mpm \end{bmatrix} \right\} \end{bmatrix}$$

Notice that the formulation of second position placement crucially relies on the use of reified morphs. The current analysis could not be formulated without further stipulations in the a-morphous framework of Crysmann and Bonami (2012): under the set of assumptions that this previous framework shares with other amorphous approaches (Anderson, 1992; Stump, 2001), the rule applying in the most peripheral order need not be a rule that actually introduces an exponent.

The final ingredient to the analysis of Sorani Kurdish person marking is the organisation of the hierarchy of morpholexical rules in conjunctive dimensions and disjunctive types (Koenig, 1999). Specifically, we assume a cross-classification in two dimensions: MORPHOTACTICS is responsible for the placement of morphs whose phonology is specified in EXPONENCE. Crysmann and Bonami (2012) shows how such a setup allows for the seamless analysis of various variable morphotactic phenomena, including positional disambiguation of person markers in Swahili, conditioned reordering in Fula and Swahili, and mobile stems in Italian. In the present case, the distinction of the two dimensions is crucial to the account of the variable placement of mobile person markers in second or final position. As Figure 3 illustrates, types in the EXPONENCE dimension enumerate the different shapes of MPMs while types in the MORPHOTACTICS list the available positioning strategies documented in Table 4 and examples (18), linking them to appropriate morphosyntactic conditions. As in Koenig (1999), individual combinations of types in conjunctive dimensions need not be listed but can be deduced by so-called online type construction: this is illustrated by the two rules at the bottom of figure 3 corresponding to second position and final placement of the 3PL marker *-jân*. Wherever the distinction between dimension does not allow such factorisation, economy of description is ensured by pre-linking the appropriate type to both dimensions, as is the case here for the rule introducing the 3SG marker *-î*, which uniformly linearises in final position (18a).

5 Conclusion

In this paper we have argued on the basis of complex morphotactic systems for a new model of realisational morphology that is characterised by two central properties: first, an information-based view of morphological completeness and coherence that crucially relies on a distinction of expression (MUD) and conditioning (MORSYN), enabling us to dispense with stipulated rule blocks altogether and to extend considerably the scope of Pāṇini's Principle. Second, by moving positional indexing from the rule system into morphological representations, we were able to provide a straightforward account of second position affixes within a much more constrained theory of inflectional morphology which denies morpholexical rules access to the full derivation history, permitting only reference to pivotal positions like that of the stem (for Italian; Crysmann and Bonami, 2012) and the left edge (for Sorani Kurdish).

A The representation of position class: some further refinements

In the body of this paper, we have assumed, without any further discussion, that position classes can be implemented by means of an integer-valued PC feature, together with a global ordering constraint, requiring morphs to be positioned in strictly ascending order. In the context of an HPSG grammar, we may represent natural numbers as lists: 0 is represented by the empty list, and for every number n, $n+1$ is represented by a list extending the list representing n by one element. We can then impose the necessary constraints on morph lists by means of the following type declarations:

(22) a. $\textit{morph-list} := \textit{list} \wedge \begin{bmatrix} \text{PREV} & \textit{list} \end{bmatrix}.$

b. $\textit{ne-morph-list} := \textit{morph-list} \wedge \textit{ne-list} \wedge \begin{bmatrix} \text{FIRST} & \begin{bmatrix} \text{PC} & \boxed{2}\left(\boxed{1} \oplus \textit{ne-list}\right) \end{bmatrix} \\ \text{PREV} & \boxed{1} \\ \text{REST} & \begin{bmatrix} \textit{morph-list} \\ \text{PREV} \quad \boxed{2} \end{bmatrix} \end{bmatrix}.$

c. $\textit{e-morph-list} := \textit{morph-list} \wedge \textit{e-list}.$

The formalisation in (22) captures the strict ordering property of morph lists, including slot competition, by means of a local type constraint: as stated in the recursive definition of a non-empty morph list in (22b), the PC feature of any morph list element must be longer than the value of the PREV feature, which represents the PC feature of the previous list item. The length of the current element's list in turn is registered on the PREV value of the list remainder, making the current position class index accessible to the next list element, if any.

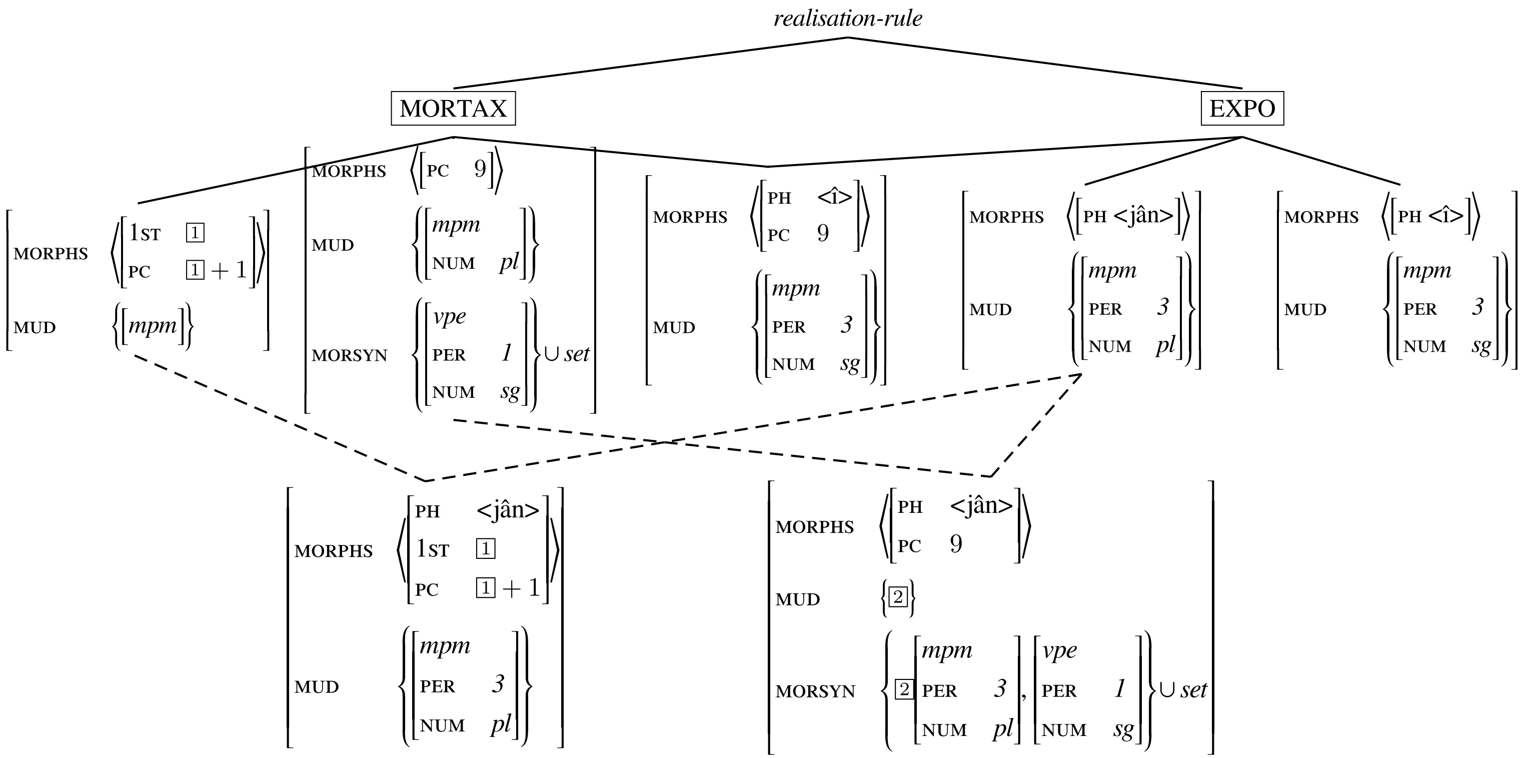

Figure 3: Realisation rules for Sorani Kurdish person markers

While this formalisation is certainly sufficient to deal with the data discussed in this paper, we shall nevertheless propose two refinements to this baseline representation: first, we shall generalise the representation of individual position classes to contiguous spans of position classes, providing a more sound approach to portmanteau morphs, and second, we shall introduce a distinction between distance and direction, thereby facilitating the treatment of ambifixals.

The first extension is rather trivial: instead of representing position classes by means of a single list, all it takes for an implementation of spans is to distinguish between start and end positions, representable as two separate list values. We therefore propose (23) as a replacement for (22b):

(23) $$\textit{ne-morph-list} := \textit{morph-list} \wedge \textit{ne-list} \wedge \begin{bmatrix} \text{FIRST} & \begin{bmatrix} \text{PC} & \begin{bmatrix} \text{FROM} & \boxed{1}\left(\boxed{0} \oplus \textit{list}\right) \\ \text{TO} & \boxed{2}\left(\boxed{1} \oplus \textit{ne-list}\right) \end{bmatrix} \end{bmatrix} \\ \text{PREV} & \boxed{0} \\ \text{REST} & \begin{bmatrix} \textit{morph-list} \\ \text{PREV} \;\; \boxed{2} \end{bmatrix} \end{bmatrix}.$$

With this first refinement in place, we can assign a positional index to portmanteau position classes, as e.g. in Swahili , without running into arbitrary decisions. Furthermore, a representation of position class built on spans provides a sound basis for slot-based competition that involves a contiguous set of positions, as witnessed, e.g., in Nimboran (Inkelas, 1993).

The second refinement we shall propose concerns the representation of ambifixal morphs, that is, morphs are found at a fixed distance from the stem (in terms of position classes), but systematically alternate between a prefixal and a suffixal realisation. Stump (1993) argues that Swahili relative markers and Fula tense and subject agreement are such ambifixals: the main intuition to be captured here is that ambifixals are found at a fixed distance from the stem (in terms of position classes), systematically alternating between a prefixal and a suffixal realisation.

The required level of abstraction from absolute to stem-relative position class can be achieved quite straightforwardly by means of an auxiliary feature DIST that is related to absolute position class indices by the following constraints on prefixal and suffixal morphs:

(24) $$\textit{pref} := \textit{aff} \wedge \begin{bmatrix} \text{PC} & \begin{bmatrix} \text{TO} & \boxed{1} \end{bmatrix} \\ \text{DIST} & \boxed{2} \\ \text{STM} & \begin{bmatrix} \text{FROM} & \boxed{1} \oplus \boxed{2} \end{bmatrix} \end{bmatrix}. \quad \textit{suff} := \textit{aff} \wedge \begin{bmatrix} \text{PC} & \begin{bmatrix} \text{FROM} & \boxed{1} \oplus \boxed{2} \end{bmatrix} \\ \text{DIST} & \boxed{2} \\ \text{STM} & \begin{bmatrix} \text{TO} & \boxed{1} \end{bmatrix} \end{bmatrix}.$$

Note that the relative order of prefixes and suffixes on morph lists follows directly from the two constraints above, together with the strict ordering of position class indices imposed by (22) and (23).

In sum, using a single indexing scheme, keyed to certain pivotal positions, we are able to capture a wide range of patterns of variation in position class system, facilitating abstraction of common properties by means of underspecification.

References

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark and Xu, Zheng. 2010. A realization optimality-theoretic approach to affix order. *Morphology* 20, 388–411.

Bickel, Balthasar, Banjade, Goma, Gaenzle, Martin, Lieven, Elena, Paudya, Netra Prasad, Rai, Ichchha Purna, Manoj, Rai, Rai, Novel Kishore and Stoll, Sabine. 2007. Free prefix ordering in Chintang. *Language* 83, 43–73.

Bonami, Olivier and Samvelian, Pollet. 2008. Sorani Kurdish person markers and the typology of agreement, présentation au *Thirteenth International Morphology Meeting*, Vienne.

Brown, Dunstan and Hippisley, Andrew. 2012. *Network Morphology: a defaults-based theory of word structure*. Cambridge: Cambridge University Press.

Bybee, Joan L. 1985. *Morphology*. Amsterdam: John Benjamins.

Carstairs, Andrew. 1987. *Allomorphy in Inflection*. Beckenham, Kent: Croom Helm.

Crysmann, Berthold. 2002. *Constraint-based Coanalysis*. Ph. D.thesis, Universität des Saarlandes.

Crysmann, Berthold and Bonami, Olivier. 2012. Establishing order in type-based realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2012*, pages 123–143, Stanford: CSLI publications.

Harris, Alice C. 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

Inkelas, Sharon. 1993. Nimboran Position Class Morphology. *Natural Language and Linguistic Theory* 11, 529–624.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Lacroix, René. 2009. *Description du dialecte laze d'Arhavi*. Ph. D.thesis, Université Lumière Lyon 2.

Luís, Ana and Spencer, Andrew. 2005. Paradigm Function Account of 'mesoclisis' in European Portuguese. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2004*, pages 79–154, Dordrecht: Kluwer.

Luutonen, Jorma. 1997. *The variation of morpheme order in Mari declension*. Helsinki: Suomalais-ugrilainen Seura.

Matthews, P. H. 1972. *Inflectional Morphology. A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Monachesi, Paola. 1999. *A lexical approach to Italian cliticization*. Stanford: CSLI Publications.

Nevis, Joel A. and Joseph, Brian D. 1992. Wackernagel affixes: Evidence from Balto-Slavic. In Geert Booij and Jaap van Marle (eds.), *Yearbook of morphology 1992*, pages 93–111, Dordrecht: Kluwer.

Nordlinger, Rachel. 2010. Verbal morphology in Murrinh-Patha: Evidence for templates. *Morphology* 20, 321–341.

Noyer, Rolf. 1992. *Features, positions and affixes in autonomous morphological structure*. Ph. D.thesis, MIT.

Rice, Keren. 2000. *Morpheme Order and Semantic Scope. Word Formation in the Athabaskan Verb*. Cambridge: Cambridge University Press.

Riese, Timothy, Bradley, Jeremy, Yakimova, Emma and Krylova, Galina. 2010. *Онгай марий йылме: A Comprehensive Introduction to the Mari Language*. Department of Finno-Ugric Studies, University of Vienna.

Samvelian, Pollet. 2007. What Sorani Kurdish Absolute Prepositions Tell Us about Cliticization. In *Texas Linguistics Society IX*, pages 265–285.

Stump, Gregory T. 1993. Position classes and morphological theory. In Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1992*, pages 129–180, Kluwer.

Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2012a. Overriding default orderings in inflectional morphology, paper presented at the Conference on Defaults in Morphological Theory, University of Kentucky, Lexington, May 21-22, 2012.

Stump, Gregory T. 2012b. Variable affix order and the amorphousness hypothesis, paper presented at the First American International Morphology Meeting (AIMM), University of Massachusetts, Amherst, September 22, 2012.

Walther, Géraldine. 2012. Fitting into morphological structure : accounting for Sorani Kurdish endoclitics. In *Proceedings of MMM8*, pages 299–321.