

## Relating Morphology to Syntax

Louisa Sadler and Rachel Nordlinger

### 1 Introduction

Relatively little attention in theoretical work in LFG has focussed on the nature of the interface between morphology and syntax, or indeed on the role of morphology proper.<sup>1,2</sup> While the contribution of morphology to the definition of f-structures is firmly established, and the separation of external structures by the principle of lexical integrity is the backbone of LFG's lexicalist outlook, the internal operation of the morphological component, and how words come to contribute the relevant f-descriptions have not generally been at the forefront of theoretical work.

From a syntactic point of view this is unsurprising, for in the general case nothing much hangs on precisely how matters internal to inflectional morphology are dealt with. In the typical case “pieces” of inflectional morphology contribute information to the f-structure of the word itself (e.g. (↑ TNS) = PAST) or to the f-structure of a dependent of the word (e.g. (↑ SUBJ PRED) = ‘PRO’), and thus the syntactic contributions of discrete “pieces” of inflectional morphology do not interact in any complicated way. This means that a simple, word-syntactic, incremental view may be taken of the syntactic contribution of inflectional morphological processes such as affixation, in which f-descriptions are associated directly with morphological forms or features.

However, more complex data makes it evident that the simple incremental view can be problematic. Difficulties may arise in several domains. For example, there is a considerable body of morphological data which suggests that an incremental approach is insufficient or inappropriate on purely morphology-internal grounds of the adequacy of morphological description, although we will not be concerned with such evidence in this paper (see Stump 2001 and Spencer this volume for more discussion). Additionally, there is a further type of complex data where the complexity relates to the interaction between the syntax and the morphology, more particularly, the interaction between the pieces of syntactic information which are encoded morphologically. Such data poses fundamental issues for the mapping between syntax and morphology, showing that the simple view, under which all that is required is to (incrementally) associate syntactic information with morphological features (or forms), is incorrect.

---

<sup>1</sup>Thanks for discussion of relevant material to Mary Dalrymple and Andrew Spencer. We are also grateful to Ryo Otuguro and two anonymous reviewers for comments on this paper. Remaining errors are of course solely our own responsibility. Sadler is grateful to the University of Essex for a period of sabbatical leave during which this work was completed, and Nordlinger for the support of the Australian Research Council, grant F9930026, held at the University of Melbourne.

<sup>2</sup>Existing work touching on, or having consequences for, these issues includes Ackerman 1990, Sadler 1997, Sells 2000 and this volume, and Kaplan and Butt 2002, amongst others, and in particular the work on verbal periphrasis including Butt et al 1996, Börjars et al 1997, Sadler and Spencer 2001 and Frank and Zaenen 2002.

This paper is concerned with data of this sort, and in particular with the phenomenon of case stacking, which shows in a clear fashion both that morphological structure *does* matter, and that there are interesting issues to be addressed in the morphology-syntax interface.

The phenomenon of case stacking, dramatically illustrated in (1) from the Australian language Martuthunira, demonstrates in an extreme form the role that morphology can have in building and constraining syntactic structures.

- (1) *Ngayu nhawu-lha ngurnu tharnta-a mirtily-marta-a*  
 I saw-PST that(ACC) euro-ACC joey-PROP-ACC  
*thara-ngka-marta-a.*  
 pouch-LOC-PROP-ACC  
 ‘I saw the euro with a joey in (its) pouch.’ (Martuthunira, Dench 1995a:60, (3.15))

In this example the most deeply embedded nominal ‘pouch’ carries three case markers, each one relating to a successively higher syntactic relationship. First ‘pouch’ is inflected with the locative case marking the f-structure function of ‘pouch’, then with the proprietive case indicating that the locative nominal is embedded within the proprietive NP ‘joey in (its) pouch’, and finally with the accusative case, marking the whole proprietive NP as being contained within the object NP. Case stacking data such as that in (1) poses some interesting challenges for morphological description because it demonstrates that morphosyntactic features may be iterated. More importantly, it casts light on the nature of the interface between syntax (f-structure) and morphology because capturing these data necessitates a complex mapping between morphological sequencing and syntactic structure.

In earlier work, Nordlinger (1998) provides an incremental, morpheme-based account of these data in LFG. The morphology (or lexical component) constructs fully inflected wordforms with multiple case markers — for this Nordlinger adopts an essentially word-syntactic approach to inflectional morphology. Functional (f-structure) information is associated directly with morphemes (which are conventionally thought of as listed as (sub-)lexical entries). The role of morphology in defining or constraining the larger syntactic environment within which the word appears is straightforwardly captured by the use of so-called *inside-out* statements. For example, on a morphemic approach the Martuthunira accusative case marker *-yu* can itself be associated with a functional description which states that it specifies or defines an OBJ function. This model of constructive morphology has been widely adopted in recent morphosyntactic research in LFG (Nordlinger 1998, Barron 1998, Sadler 1998, Sells 2000, Lee 1999, Sharma 1999, Nordlinger 2000, Nordlinger and Sadler 2000, O’Connor 2002, Ørsnes 2002). To accommodate case stacking, Nordlinger formulates a combinatorial principle (the Principle of Morphological Composition (PMC)) which correctly constrains the interaction of f-structure information associated with different morphemes in the inflected word. However an important issue which arises in this connection

is that the PMC is not formalized within the LFG description language in that work.<sup>3</sup> As should become clear as we proceed, one reason why the mapping between morphemes (or morphological features) and syntactic functions is problematic to formalize on a word syntactic view is precisely because the incremental morphemic view under which affixes are added hierarchically by means of a binary  $X \rightarrow X$  Aff rule provides an inappropriate structure for the felicitous combination of f-descriptions.

The purpose of this paper is to explore more fully the mapping between morphological and syntactic descriptions, with primary reference to the case stacking data. In particular we aim to show how the insights of Nordlinger's (1998) Principle of Morphological Composition can be incorporated into a model of the morphology/syntax interface which is fully compatible with the description language of LFG, but on the basis of different assumptions about the nature of the morphological representation itself. We will not be concerned here with actual morphological forms, nor with the theory of inflectional morphology which generates the inflected word forms. In specifying the mapping between morphology and syntax in this paper our starting point will be the sorts of structured representations delivered by the inferential-realizational account of the morphology of case-stacking proposed in Sadler and Nordlinger (to appear).<sup>4</sup>

The rest of this paper is structured as follows. Section 2 provides the background necessary to understand the problem we focus upon. It begins with a brief review of the relevant case stacking data and then illustrates the constructive morphology approach to this data which our analysis builds on. The section concludes with an evaluation of the Principle of Morphological Composition. With this background in place, section 3 outlines our proposal for the interface between morphological structures and f-descriptions and section 4 concludes.

---

<sup>3</sup>Note however that the appendix to Nordlinger (1998) suggests an approach to restating the PMC with standard LFG tools which prefigures in some respects the approach to formalization which we adopt here.

<sup>4</sup>Sadler and Nordlinger (to appear) propose a morphological analysis in the framework of Paradigm Function Morphology (PFM), which has several advantages over the morpheme-based account of Nordlinger (1998). In particular, it is quite straightforward in PFM to set up the morphology to appropriately constrain the interaction of different case functions (e.g. the relational, modal, associating and complementizing case functions of Kayardild (Evans 1995a)), whereas in the original work by Nordlinger, the morphology overgenerates on this front, with the syntax effectively playing a filtering role. Additionally, the separation of function from exponence in a realizational framework permits various exceptional forms, such as portmanteau affixes and case substitutions, to be correctly treated. See Sadler and Nordlinger (to appear) for further discussion.

## 2 Composition and F-Descriptions

### 2.1 Case stacking

The following is a straightforward example of adnominal multiple case in the Australian language Thalanyji. Here the possessor is coded by the dative case marker *-ku* DAT, and also takes the case of the head that it modifies (ACC).

- (2) *kupuju-lu kaparla-nha yanga-lkin wartirra-ku-nha*  
 child-ERG dog-ACC chase-PRES woman-DAT-ACC  
 ‘The child chases the woman’s dog.’ (Thalanyji, Austin 1995:372, (22))

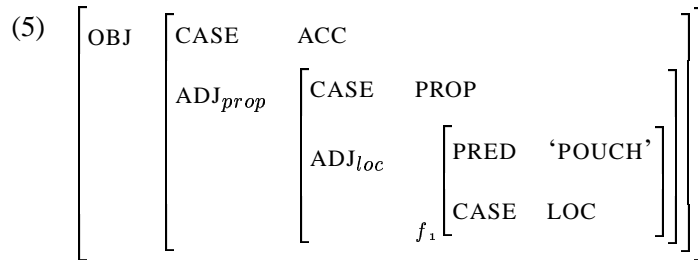
The fully inflected word *wartirra-ku-nha* ‘woman-DAT-ACC’ projects the functional information shown in (3), in which  $f_1$  is the f-structure of the nominal itself, and contains the dative case feature, while the accusative case feature belongs in a higher f-structure (namely, that of the head noun ‘dog’).

- (3) 
$$\left[ \begin{array}{cc} \text{CASE} & \text{ACC} \\ \text{POSS} & \left[ \begin{array}{cc} \text{PRED} & \text{‘WOMAN’} \\ \text{CASE} & \text{DAT} \end{array} \right] \end{array} \right]$$

A more complex case stacking example is the following from Martuthunira, repeated from (1) above.

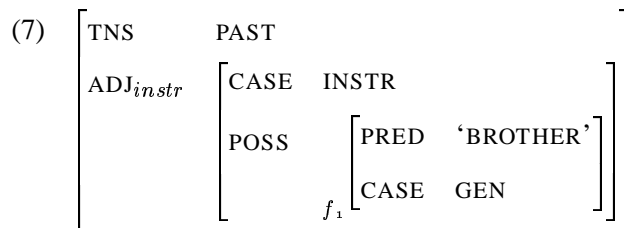
- (4) *Ngayu nhawu-lha ngurnu tharnta-a mirtily-marta-a*  
 I saw-PST that(ACC) euro-ACC joey-PROP-ACC  
*thara-ngka-marta-a.*  
 pouch-LOC-PROP-ACC  
 ‘I saw the euro with a joey in (its) pouch.’

The information projected from the single inflected nominal *thara-ngka-marta-a* ‘pouch-LOC-PROP-ACC’ is shown diagrammatically in (5) — once again, the f-structure of the nominal itself is  $f_1$ . In this example, the innermost (locative) case marks the locative adjunct ‘pouch’. The proprietive case signals the proprietive adjunct relation between ‘joey’ and ‘euro’, and the outermost (accusative) case signals the relation between ‘euro’ and the verb (the object relation).



The case system of the Tangkic language Kayardild (see Evans 1995a) is still more complex, additionally permitting the use of case markers in modal function (in which case morphology partially specifies temporal and modal information at the level of the clause) and case markers in complementizing function, where case markers are used to mark interclausal relations on complementized clauses (Dench and Evans 1988, Evans 1995a). The word *thabuju-karra-nguni-na* in (6) illustrates the combination of two case markers in core function with a case marker in modal function. The modal ablative case (M.ABL) marks the clause as having past tense. The instrumental case marks ‘brother’ as belonging to an instrumental argument of the verb and the genitive case marks ‘brother’ as the (adnominal) possessor argument within the instrumental NP. The information projected from the nominal *thabuju-karra-nguni-na* is shown in (7)— again, the f-structure of the nominal itself is  $f_1$ .

- (6) *Ngada yalawu-jarra yakuri-na thabuju-karra-nguni-na mijil-nguni-na.*  
 I catch-PST fish-M.ABL brother-GEN-INST-M.ABL net-INST-M.ABL  
 ‘I caught the fish with brother’s net.’ (Kayardild, Evans 1995b: 400, (10))



As a further complication to the data, note that number marking may be interleaved with case marking in these languages, with each instance of number marking modifying a different referent according to its position in the morphological structure, as shown in the Kayardild example (8), in which the ablative case marks the possessor function.

- (8) *maku-yarr-nurru-naba-walad*  
 woman-DU-ASSOC-ABL-MANY(NOM)  
 ‘the many belonging to (those) having two wives’ (Kayardild, Evans 1995a:123)

Similarly, in the following Martuthunira noun phrase, the pronominal stem is SG, the genitive case marks the possessor function and the subsequent dual number marking is interpreted with respect to the possessed noun *pawulu-* ‘child’.

- (9) *nganaju-wu-tharra-a pawulu-tharra-a*  
 1SG.OBL-GEN-DU-ACC child-DU-ACC  
 ‘my two children’ (Dench 1995a:95, (4.154))

- (10) 
$$\left[ \begin{array}{cc} \text{NUM} & \text{DU} \\ \text{CASE} & \text{ACC} \\ \text{POSS} & \left[ \begin{array}{cc} \text{CASE} & \text{GEN} \\ \text{PRED} & \text{'PRO'} \\ \text{NUM} & \text{SG} \\ \text{PERS} & 1 \end{array} \right] \end{array} \right]_{f_1}$$

The fact that number marking interacts with the stacking of case markers in this manner is significant because it demonstrates that more is at issue here than a simple “quirk” of the case system of these languages. Rather it is evident that the morphological structure itself is complex and that the successive levels of case marking define syntactic structures which are referenced by other f-structure descriptions expressed word-internally, thereby providing additional support for the constructive view itself.

## 2.2 Constructive Morphology: Inside-Out Constraints

Such case stacking clearly demonstrates the fundamental role that morphology can play in encoding complex syntactic relations. In order to account for such data Nordlinger (1998) develops the model of constructive case within LFG whereby morphological constituents/processes may actively define properties of their clausal environment independently of syntax (see also Simpson 1983, 1991 and Andrews 1996). The model of constructive case consists essentially of two distinct ideas.

The first of these is the use of inside-out constraints (e.g. Halvorsen and Kaplan 1988, Dalrymple 1993, see also Andrews 1996:41-43) associated with the lexical elements or morphological processes to enable nominal constituents to define the larger syntactic (f-structure) context in which they are embedded.<sup>5</sup> In this way, case-marked nominals can specify the grammatical function of the higher clause of which their f-structure is the value. Thus the f-structure information associated with accusative

<sup>5</sup>Inside-out function application is well-established in LFG through work in such areas as quantifier scope (Halvorsen and Kaplan 1988), anaphoric binding (Dalrymple 1993), internally-headed relative clauses (Culy 1990), Russian genitive of negation (King 1995), Urdu case (Butt 1995), and topicalization (Bresnan 2001).

case in an example such as the Martuthunira (4) is as in (11), and the accusative-case nominal (e.g. ‘euro-ACC’) projects the f-structure in (12).

$$(11) \quad \text{ACC: } (\uparrow \text{ CASE}) = \text{ACC} \\ (\text{OBJ } \uparrow)$$

$$(12) \quad \left[ \text{OBJ } \left[ \text{PREL } \text{‘EURO’} \right] \right]_{f_x} \\ \left[ \text{CASE } \text{ACC} \right]_{f_y}$$

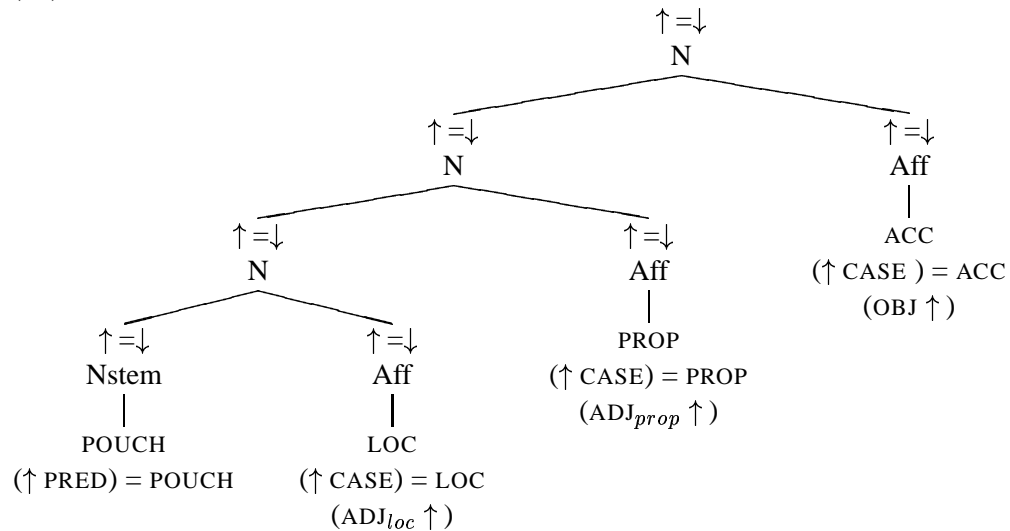
By virtue of the inside-out designator ( $\text{OBJ } \uparrow$ ), the information associated with the accusative case constructs a higher f-structure ( $f_x$ ) which contains an OBJ to which the f-structure associated with the case-inflected nominal itself ( $f_y$ ) belongs. Thus, on this analysis, a nominal inserted into the syntax already defines its grammatical function by virtue of the case marker attached to it.

While the use of the inside-out constraints enables nominals to construct information about the higher f-structure in which they are embedded (e.g. by specifying a grammatical function for it), this alone does not provide an analysis of the case stacking data presented above. The constructive case model also contains a second component to compose the information contributed by multiple morphological elements, for each case marker contributes information about a successively higher f-structure. This is the Principle of Morphological Composition, which we discuss in the following subsection.

### 2.3 Interfacing Stacking Morphology with Syntax

On a standard LFG view of how the f-description associated with an inflectionally complex word comes about, the syntactic information associated with each morphological element is simply conjoined. For a word with multiple case markers, such as *thara-ngka-marta-a* ‘pouch-LOC-PROP-ACC’ in (4), this essentially corresponds to positing the following sub-lexical annotated tree, in which each case marker introduces a CASE value and makes reference to a (different) grammatical function:

(13)



But of course this will not give the desired results because it fails to embed the f-structures in the appropriate manner (and as a consequence it also assigns multiple inconsistent CASE values to the f-structure denoted throughout by  $\uparrow$ ). Nordlinger (1998) observes that, informally speaking, each successive affix takes the outer f-structure (call it  $f_x$ ) described by the previous affix and defines some properties of both it and the higher f-structure which immediately contains  $f_x$ . If the process of affixation is constrained to have this syntactic consequence, then the iconic effects exemplified in section 2.1 above are accounted for. This insight is captured in the Principle of Morphological Composition, which composes the information associated with successive affixes. According to the PMC, the f-structure information the affix actually defines depends on the information associated with the preceding morphological element (which could be the root or a more deeply embedded affix). The idea is that context-sensitive substitutions to the f-descriptions in (sub-)lexical entries are carried out: every occurrence of  $\uparrow$  in the lexical information associated with an affix is substituted with any expression of the form  $(GF^n \uparrow)$  on the preceding morphological element. Nordlinger (1998) formulates this principle as follows:

(14) **Principle of Morphological Composition:** Where  $x$  is a string of attributes:

$$\begin{array}{ccc}
 \text{Stem} & \text{Aff} & \\
 \textcircled{(GF^n \uparrow)} & ((GF^m \uparrow)) & x \\
 \hline
 \text{Stem} & \text{Aff} & \\
 (GF^n \uparrow) & ((GF^m (GF^n \uparrow))) & x
 \end{array}
 \implies$$

In the case where the preceding morphological element is annotated simply with outside-in equations, such as  $(\uparrow \text{ TNS}) = \text{ PAST}$ , the intention is that substitution of  $\uparrow$  by  $\uparrow$  will apply vacuously, while on the other hand if the previous element is annotated  $(\text{ADJ } \uparrow)$  then every occurrence of  $\uparrow$  on the current affix is replaced by  $(\text{ADJ } \uparrow)$ . In this way, each affix defines information pertaining to a successively larger, con-







$$(20) \left[ \begin{array}{ll} \text{CASE} & \text{ERG} \\ \text{NUM} & \text{DU} \\ \text{ADJ} & \left[ \text{PRED} \quad \text{'SMALL'} \right] \end{array} \right]$$

The information associated with the (sub-)lexical entries is as shown in the first column of (21) and the f-descriptions derived by the operation of the PMC in the second column:

|      | Sub-Lexical Entry                     | Entry Derived In Context      |
|------|---------------------------------------|-------------------------------|
| (21) | WITA<br>(↑ PRED) = 'small'<br>(ADJ ↑) | (↑ PRED) = 'small'<br>(ADJ ↑) |
|      | ERG<br>(↑ CASE) = ERG                 | ((ADJ ↑) CASE) = ERG          |
|      | DU<br>(↑ NUM) = DU                    | ((ADJ ↑) NUM) = DU            |

In fact, the (ADJ ↑) annotation associated with nominal stems such as *wita* is optional. General LFG principles of completeness and coherence will ensure that grammatical f-structures only result if the ADJ function is present for a modifier use and absent when the nominal functions as the NP head.

Similarly, in (19) the possessive pronominal root is inflected for the case and number of the possessed element, which is the head of the containing f-structure, while the pronominal root itself introduces the POSS function. The structure defined by the fully inflected word *nangi-marnda-rna* is shown in (23).

$$(22) \quad \begin{array}{ll} \textit{nangi}: & (\text{POSS } \uparrow) \\ & (\uparrow \text{PRED}) = \text{'PRO'} \quad (\uparrow \text{GEND}) = \text{'MASC'} \\ & (\uparrow \text{PER}) = \text{'3'} \quad (\uparrow \text{NUM}) = \text{'SG'} \end{array}$$

$$(23) \left[ \begin{array}{ll} \text{CASE} & \text{ACC} \\ \text{NUM} & \text{PL} \\ \text{POSS} & \left[ \begin{array}{ll} \text{PRED} & \text{'PRO'} \\ \text{PER} & \text{3} \\ \text{GEND} & \text{MASC} \\ \text{NUM} & \text{SG} \end{array} \right] \end{array} \right]$$

Notice that once the word formation component is set up to provide the right forms, the constructive morphology approach exemplified here immediately accounts, without further modification, for forms such as this which carry two number and gender

values – one for the possessor and one agreeing with that of the possessed element. The Principle of Morphological Composition, in this case in conjunction with the assumption that the stem introduces the POSS function, ensures that the values SG and MASC are in the f-structure of the possessor and the values PL and FEM in the f-structure of the governing nominal ‘daughter’.

## 2.5 Evaluating the PMC

The PMC straightforwardly captures the intuition that as they stack, the (case and number) affixes contribute information to the f-structure defined by the morphological structure to which they attach. This places a strong constraint on the relationship between syntax and morphology, imposing a sort of isomorphism: morphological structure and syntactic structure are required to match in the appropriate sense. Because the principle essentially embeds the syntactic information associated with the previous (more deeply embedded) affix into the syntactic information associated with the next (higher) affix, it automatically accounts for the observed iconic behaviour of case stacking.

Despite this, the principle itself is not without problems. In particular, the statement of this principle does not fall within the mathematics which underpins the LFG formalism. As we have seen, the operation of the PMC entails the *substitution of paths* in the f-descriptions associated with affixes. Although the formulation of the principle in (14) might give the impression that these substitutions are local to a sub-tree of depth one, note that this is not actually the case — the substitution is not between sisters in a local subtree. In fact, a crucial aspect of the principle is that the relevant substitutions are really performed at the level of the information lexically associated with the affixes, and not at the level of the derived word. As shown above, the word (and indeed the stem at each level of recursion within the word structure) is annotated  $\uparrow = \downarrow$ . This is crucial to ensuring that the word itself interacts correctly with the f-descriptions associated with the other nodes in the c-structure tree. What the principle ensures is that the information associated in the lexicon with an affix is modified by substituting for the  $\uparrow$  designation on the affix whatever inside-out path is eventually associated with the preceding affix, once any substitutions at that level have been performed. This requires a form of pattern matching, which is not supported within the LFG formalism.

Because the approach to word internal structure in Nordlinger 1998 is basically morphemic, the PMC is formulated with reference to trees which reflect morphemic structure. Once we move away from a morpheme-based morphology, however, we are able to consider different structurings of morphological information which are not tied to exponence, and as a consequence the effect of the principle of morphological composition can be captured directly without the need for any path substitutions. In the rest of this paper, we take as our starting point the recent morphological treatment of the Kayardild case stacking data in Sadler and Nordlinger (to appear) and show how

the mapping to syntax can be directly encoded on the basis of these morphological structures.

### 3 The Morphology-Syntax Interface

We noted at the start of this paper that the interface between (inflectional) morphology and syntax has received relatively little attention to date in LFG, the standard view being a lexicalist, incremental one in which *f*-descriptions are associated with elements of the morphology (for example, affixes or morphological features) and then combined straightforwardly by identifying word-internal instances of  $\uparrow$ . In very many cases, of course, this is unproblematic, for the simple reason that the information expressed by a word is typically quite local to the *f*-structure of that word.

Conceptually, the interface between lexical representations and syntactic (*f*-) descriptions involves two distinct aspects. The first of these is the specification of a mapping between syntactic information and corresponding elements in the morphological domain, that is, morphological features, and lexemes or roots. In very many cases, of course, this mapping is highly transparent, to such an extent that it is often tacitly assumed that the syntax and the morphology involve one and the same set of morphosyntactic properties, but in fact there are good reasons to keep these features distinct. For one thing, there are clearly morphology-internal features (such as conjugation class), which have no place in syntactic representations of any sort, and for another, there are well-known cases of mismatch between morphological and syntactic features, for example where an element is morphologically a member of category A but syntactically a member of category B (see Spencer to appear). The second aspect of the interface between morphology and syntax is the specification of how the syntactic information associated with the morphological “parts” is to be combined. Here too, within an LFG context, simple concatenation of *f*-descriptions associated with morphological “parts” (with identification of instances of  $\uparrow$ ) is generally appropriate. However, as we have seen, this is not the case for the case stacking data, in which the functional information associated with morphological “parts” interacts in a more complex way. In the current section, we show how this interaction can be simply captured in the interface between a realizational morphology and the syntax.

#### 3.1 Associating *F*-descriptions with *M*-features

In recent work, Sadler and Nordlinger (to appear) provide a morphological account of case stacking within Paradigm Function Morphology (PFM) (Stump 2001). In such a realizational approach, the morphological descriptions are paradigm cells, where a paradigm cell is a pair consisting of a lexeme (or root) and a well-formed fea-

ture bundle.<sup>6</sup> For example, the morphological structures associated with fully inflected wordforms in a case stacking language such as Kayardild (which has a modal case function as well as core (adnominal or relational) case functions) are as shown in (24). The structure in (24 a) is associated with the word *thabuju-karra-nguni-na* ‘brother-GEN-INST-M.ABL’ (see (6)), while the nominal *maku-yarr-nurru-nabawalad* ‘woman-DU-ASSOC-ABL-MANY(NOM)’ (see (8)) is of the more complicated form involving multiple number features, shown schematically in (24 b). These examples are given in a notation for specifying paradigm cells in PFM (in what follows we will often abbreviate  $\{Core\}$  as  $\{C\}$ ).

- (24) a. <LEX, {Case<sub>Core</sub>:W, {Case<sub>Core</sub>:Y, {Case<sub>Mod</sub>:X}}}>  
 b. <LEX, {Num:A, {Case<sub>Core</sub>:W, {Case<sub>Core</sub>:Y, {Num:X, {Case<sub>Core</sub>:Z}}}}}>  
 c. <LEX, {Num:A, {Case<sub>Core</sub>:W, {Case<sub>Core</sub>:Y, {Case<sub>Mod</sub>:X}}}}>

Sadler and Nordlinger (to appear) is concerned essentially with morphology-internal matters, rather than with the nature of the interface between morphological structures and f-descriptions, and includes only a brief sketch of a relatively simple procedure for correctly combining the syntactic information associated with each morphological feature (and the lexemic root). The current paper, on the other hand, takes as its starting point the sorts of structured morphological representations proposed in Sadler and Nordlinger (to appear) and shows how the mapping between morphological structures and the syntax can be straightforwardly specified without any extension to the LFG formalism (but with the addition of a minor notational extension), using f-descriptions directly in the morphology-syntax interface.

As noted above, the first step is to specify the syntactic information corresponding to each morphological feature. This is quite trivial, and the syntactic information is precisely that proposed in Nordlinger (1998), although of course in our realizational approach, this information is associated with morphological descriptions or features and not with morphemes. A mapping is specified between morphological A:V pairs and syntactic information, and between roots and syntactic information — this may be thought of as a lexicon or lexical transducer.<sup>7</sup> We show below some examples of (constructive) case, number and nominal stems:

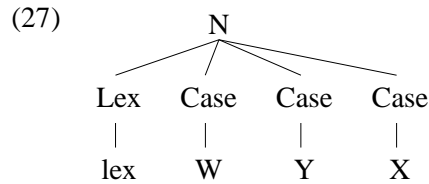
<sup>6</sup>That is, these structures are the output of morphological (inflectional) analysis, and the input to morphological (inflectional) generation. The rules of (inflectional) morphology relate well-formed paradigm cells to realizations.

<sup>7</sup>See Kaplan and Newman 1997 and Butt et al 1999 for discussion of a similar “lexicon” of morphological formatives within the XLE computational environment.

| Mfeature                    | F-description                          |
|-----------------------------|--|
| Case <sub>C</sub> :Loc      | (↑ CASE)=LOC, (ADJ <sub>loc</sub> ↑)   |
| Case <sub>C</sub> :Abl      | (↑ CASE)=ABL, (ADJ <sub>abl</sub> ↑)   |
| Case <sub>C</sub> :Prop     | (↑ CASE)=PROP, (ADJ <sub>prop</sub> ↑) |
| (25) Case <sub>C</sub> :Erg | (↑ CASE)=ERG, (SUBJ ↑)                 |
| Case <sub>C</sub> :Nom      | (↑ CASE)=NOM, (SUBJ ↑)                 |
| Case <sub>M</sub> :Abl      | (↑ TNS)=PAST                           |
| Num:Many                    | (↑ NUM) = MANY                         |
| Num:Pl                      | (↑ NUM) = PL                           |
| woman                       | (↑ PRED) = 'WOMAN'                     |

The second step is to specify how the syntactic information, that is, the f-descriptions, are to combine. The lexical form, that is, in PFM the paradigm cell corresponding to a fully inflected word, can be viewed as a simple tree structure in which each attribute is a preterminal node and each value a terminal node – this is captured in (26) where MF is a metavariable over feature labels Num, Case, and so on. For example, the paradigm cell in (24 a) would define a structure as shown schematically in (27) below.

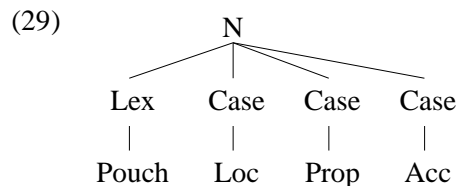
(26)  $N \rightarrow \text{Lex MF}^*$



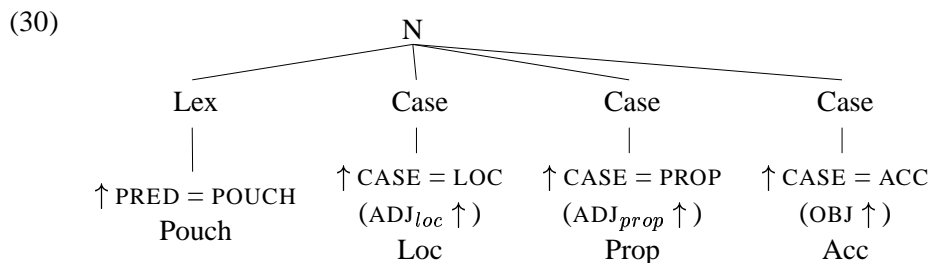
To take an example, we consider the Martuthunira nominal *thara-ngka-marta-a* ‘pouch-LOC-PROP-ACC’ from example (4). The morphological description is as follows:

(28)  $\langle \text{thara}, \{\text{Case}_C:\text{Loc}, \{\text{Case}_C:\text{Prop}, \{\text{Case}_C:\text{Acc} \}\}\} \rangle$

and the interface tree is represented as in (29) in which morphological feature values are shown as terminal nodes with initial capitalisation.



The f-descriptions are associated with the terminal nodes, giving (30):

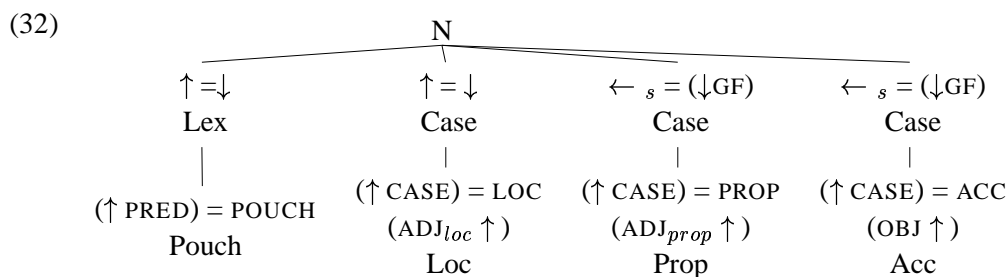


It remains to specify the annotations associated with the pre-terminal tree nodes. The generalization for a language with a constructive case feature, such as Kayardild or Martuthinira, is as follows:

- (31) Annotation Principle: For node  $n$ , if the immediately preceding left sister node is Case then annotate node  $n$  with  $\leftarrow_s = (\downarrow GF)$ , otherwise annotate with  $\uparrow = \downarrow$

The  $\leftarrow_s$  appearing here is a notational innovation which we discuss below. Note that the interface grammar does not play a filtering role, rather it is the role of the morphology proper to ensure that only well-formed feature bundles are ever produced, for example by ensuring that Number is never directly embedded under Number, and so forth.

The two operations of the interface, namely lexical look-up and (re-)parsing of the morphological structure, together with the annotation principle, provide the annotated tree-structure shown in (32).



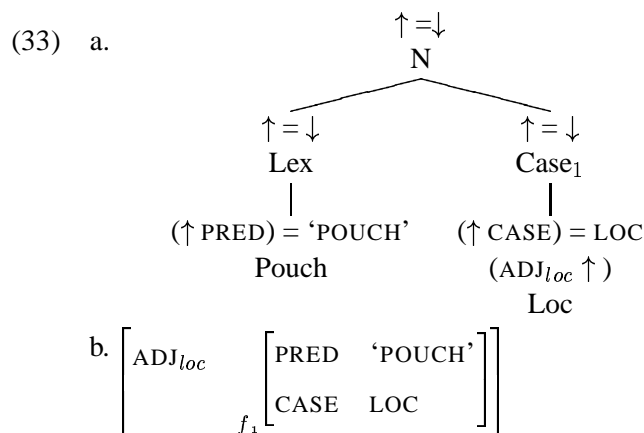
The notational innovation left arrow ( $\leftarrow_s$ ) denotes the f-structure of the immediate left sister of the node to which it is attached: that is, while  $\uparrow$  denotes the f-structure of the mother node ( $\phi(M^*)$ ) and  $\downarrow$  denotes the f-structure of the node to which it is annotated ( $\phi(*)$ ),  $\leftarrow_s$  denotes  $\phi(LS^*)$ , where  $LS$  stands for a function which



picks out an immediately preceding sister node. Note that  $\leftarrow_s$  is distinct from the  $\leftarrow$  function used in LFG to refer to the immediately containing f-structure in specifying off-path constraints (Dalrymple 1993),<sup>8</sup> and for this reason we attach an *s* subscript to the left arrow we use here,  $\leftarrow_s$ .<sup>9</sup>

### 3.2 Substitution and Simplification

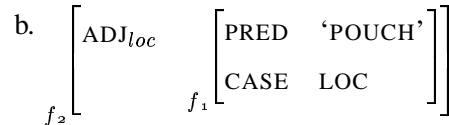
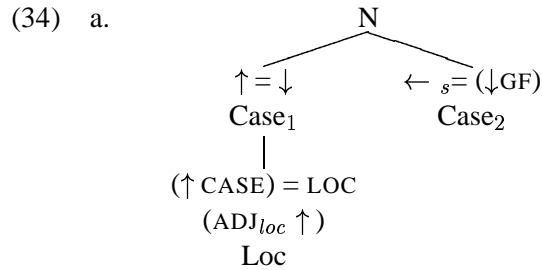
For clarity, we work through the process of constraint simplification and satisfaction for (32) step by step to illustrate what  $\leftarrow_s$  does: since this involves multiple Case nodes we label them with subscripts in the tree fragments to improve comprehensibility. The annotations associated with the tree fragment in (33 a) define the partial f-structure also shown in (33 b).



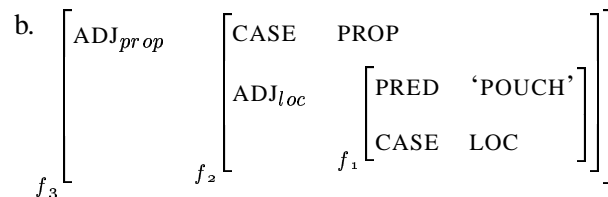
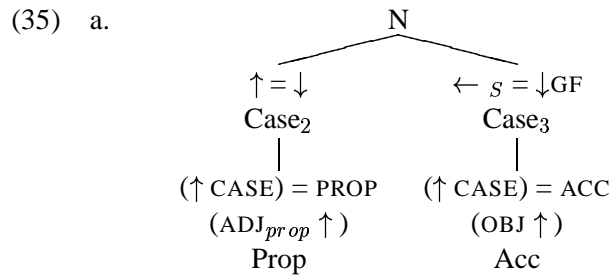
Continuing through the tree left to right, consider the tree fragment in (34 a). The annotation  $\leftarrow_s$  on the node  $\text{Case}_2$  specifies that the f-structure of the sister node ( $\text{Case}_1$ , of which the f-structure is  $f_1$ ) is the value of some GF in the f-structure ( $f_2$ ) of the node  $\text{Case}_2$ . Since  $f_1$  is the value of the path  $\text{ADJ}_{loc}$  (from above),  $\text{ADJ}_{loc}$  and GF are equated, defining the structure in (34 a).

<sup>8</sup>Off-path constraints are used especially in the statement of conditions on long distance dependencies and anaphoric dependencies. In this context,  $\leftarrow$  associated with an attribute *a* denotes the f-structure of which *a* is an attribute (see Dalrymple 2001:151).

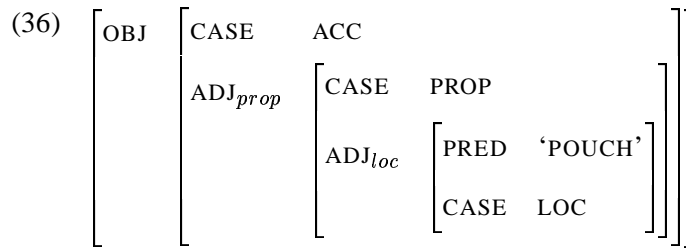
<sup>9</sup>It seems very possible that  $\leftarrow_s$  will be useful more generally in the description of syntactic phenomena. In recent computational work on projecting f-structure from chunk-based shallow trees, Frank (2003) independently proposes the addition of  $\leftarrow$  to refer to the f-structure of the (left) sister.



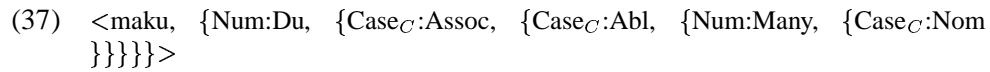
Substitutions in the rest of the f-descriptions operate in exactly the same fashion. For example, for the next tree fragment, (35 a), we know that f-structure of the Case<sub>2</sub> node is  $f_2$ . The annotations on the terminal node specify that  $f_2$  is the value of  $\text{ADJ}_{prop}$  in a larger f-structure and define  $f_2$ 's own CASE feature as PROP. The annotation  $\leftarrow_s$  on the immediate right sister node Case<sub>3</sub> specifies that the f-structure of the immediate left sister node (Case<sub>2</sub>, f-structure  $f_2$ ) is the value of some GF in the f-structure ( $f_3$ ) of the node Case<sub>3</sub>. Since  $f_2$  is the value of the path  $\text{ADJ}_{prop}$  (from the Case annotations),  $\text{ADJ}_{prop}$  and GF are equated (35 b).



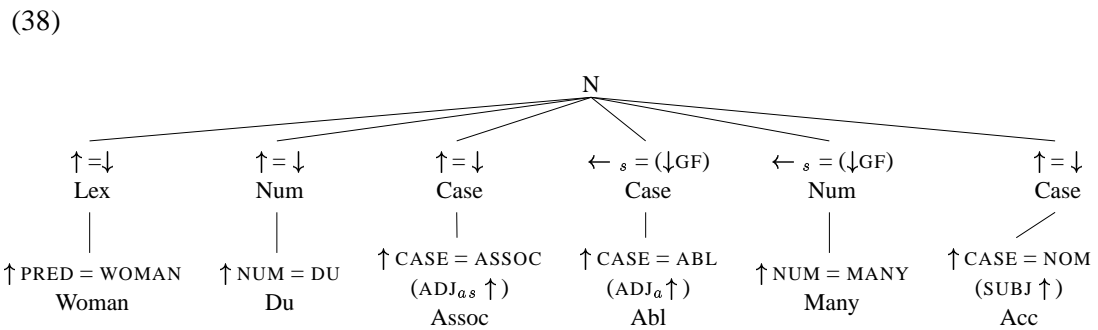
The annotations on the terminal node under Case<sub>3</sub> specify that  $f_3$  is the value of OBJ in some larger f-structure and define  $f_3$ 's own CASE feature as ACC. Thus, the f-descriptions associated with the tree in (32) are satisfied by the following f-structure in the minimal model:



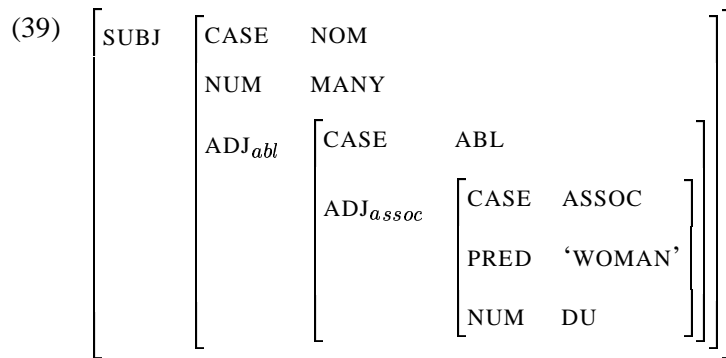
Turning now to a second example, this time involving the interaction of case and number marking, the morphological description for the Kayardild nominal in (8) is as shown in (37).



The two operations of the interface, namely lexical look-up and (re-)parsing of the paradigm cell structure provide the annotated tree-structure (38).



These f-descriptions are satisfied by the structure in (39).



### 3.3 Constructive Stems Revisited

We turn now to the treatment of constructive stems. These are the cases, exemplified by (18) and (19), in which the root itself introduces a grammatical function. This



Under the proposal made here, therefore, the f-structure projected by the fully inflected word *wita-jarra-rlu* is the f-structure of the head that *wita-* itself modifies.

As an alternative to the current proposal we might instead consider dealing with constructive stems by introducing a morphological feature into the feature bundle with no exponent. For example, for the Warlpiri adjunct *wita-jarra-rlu* we would introduce a Case feature (with no corresponding exponent) at the top of the feature bundle, which would map to an ADJ function, and for the Wambaya possessive pronoun *nangi-marnda-rna* 3.SG.M.POSS-PL-II(ACC) we would introduce a Case feature (with no corresponding exponent) which would map to a POSS function. This approach would certainly reduce the set of constructive roots to instances of constructive inflectional morphology, but it seems otherwise to be incorrect. Firstly, the functional ambiguity of nominals in a language such as Warlpiri is quite systematic and is simply not dependent on morphological case. Secondly, this approach would make the incorrect claim that those languages which have constructive roots necessarily permit case stacking — this claim is incorrect for some languages, including Wambaya. For these reasons, we do not pursue here this alternative approach.<sup>10</sup>

### 3.4 Evaluation

This section has outlined a new proposal for associating f-descriptions with morphological structures. The approach to the morphology-syntax interface involves deriving a simple flat tree structure from the structured morphological representation. The nodes of this morphological structure tree are annotated with f-descriptions in a familiar fashion, and the resultant equations simplified. The major advantage of this approach over the morpheme-based PMC of Nordlinger (1998) (and the simplified threading technique of Sadler and Nordlinger (to appear)) is that, with the exception of a modest notational extension to the language of f-descriptions, it uses the standard LFG formalism with no consequences for the formal power of the language. No pattern matching substitutions are required because the morphological tree is not based on incremental affixation in morphemic fashion.

From a linguistic point of view, however, the fundamental issue is the extent to which

<sup>10</sup>The approach that we take here to constructive stems shows some commonality with the treatment of *derivational* case morphology. It can be established that in some Australian languages some case morphology is derivational rather than inflectional, producing an inflectional stem which defines its own grammatical function (see, for example, Austin 1995, Nordlinger 1998). For example, in Wambaya, which does not permit inflectional case stacking, the cases PROP and PRIV are derivational while other cases are inflectional. In an example such as (1), the derivational morphology produces the root *gijilulunguj*, which is constructive, introducing an ADJ function.

- (1) *Yandu ngi-n                      bungmaj-buli-ja    gijilulu-nguj-nuli-ja*  
 wait 1.SG.S(PRES)-PROG old.person-DU-DAT money-PROP-DU-DAT  
 I'm waiting for the two old women with money (Nordlinger 1998:115 (41))

this proposal for the interface captures the basic intuition about constructive morphology. This intuition is the idea that (at least in these languages), morphological and syntactic structures (of the appropriate sorts) are nested or hierarchicalized isomorphically; the syntactic contribution of a particular affix “builds” on that of the (perhaps complex) stem to which it attaches. This intuition is captured in the original PMC formulation by the incremental transformation or substitution into the f-description of a higher piece of morphological structure as a function of the (input from) the f-description of the immediately lower or contained piece of morphological structure/affix. Although not formalized, the substitution is stated in such a manner as to apply invariantly to both constructive and non-constructive morphological features (that is, irrespective of whether or not the f-description contains an inside-out statement). Thus, the syntactic ramifications of a morphosyntactic feature follows purely from the f-descriptions associated with it (namely, whether or not it constructs a grammatical function), and not from the morphosyntactic structure (which remains the same irrespective).

In the current approach, however, a distinction between constructive and non-constructive features is also made in the Annotation Principle (31): the presence of a constructive feature (Case) triggers the  $\leftarrow_s = (\downarrow\text{GF})$  annotation, while a node following any other category is annotated  $\uparrow = \downarrow$ . Thus the distinction between constructive and non-constructive features is made twice: in the f-descriptions introduced on terminal and on non-terminal nodes in the mapping structure. While this might be thought to be a disadvantage, it should be noted that the present proposal, which limits the permissible annotations to  $\uparrow = \downarrow$  and  $\leftarrow_s = (\downarrow\text{GF})$ , does rule out anti-iconic relations between morphological features and f-structures, as does Nordlinger’s original proposal.<sup>11</sup> Therefore, while the current approach may not have the generality of Nordlinger’s (1998) PMC it has the same empirical adequacy and, furthermore, is straightforwardly integrated into the standard LFG architecture.

## 4 Conclusion

In most instances, the inflectional features of words define or reflect properties of the very local context — verbs express the tense, aspect and mood features of the clauses which they head, and encode agreement properties of their dependent core arguments, and nouns, adjectives and determiners inflect for properties such as the number, gender and definiteness of the nominal f-structure which they co-define. The phenomenon of case stacking in Australian Aboriginal languages, however, shows that inflectional morphology may express syntactic information pertaining to a much larger context, with words expressing functional information pertaining to f-structures

---

<sup>11</sup>It could be argued, however, that in any case a restriction to iconic ordering should be given a functional rather than a grammatical explanation, and therefore should not form a part of the grammar. Whether or not this position is adopted, it remains a fact that the grammar must be able to accommodate the complex contributions to relational structure that the case stacking data exemplifies.

within which their own f-structures are deeply embedded. Furthermore, the association of wider syntactic information with “pieces” of inflectional morphology comes about in a highly structured manner. The importance of this data is that it provides a crucial window onto the nature of the interface between inflectional morphology and functional structure: the f-structure information associated with inflectional morphology must interact in a structured way. For these languages, associating f-descriptions with the sort of morphological structure assumed by a word-syntactic, morphemic model gives the wrong result, because the f-descriptions interact incorrectly. On the other hand, given an approach to inflectional morphology which relates structured property sets to exponents, the correct interaction between f-descriptions associated with morphological features can be obtained by representing morphological structures *in the interface* as relatively flat trees, and associating and resolving f-descriptions associated with tree nodes in the normal fashion, even for languages with complex case stacking morphology of this sort. The present proposal therefore permits us to interface a realizational approach to inflectional morphology with a standard LFG syntax in a straightforward manner.

## Bibliography

- Ackerman, F. 1990. The Morphological Blocking Principle and Oblique Pronominal Incorporation in Hungarian. In P. F. Katarzyna Dziwirek and E. Mejías-Bikandi (Eds.), *Grammatical Relations: A Cross-Theoretical Perspective*, 1–20. Stanford, CA: CSLI Publications.
- Andrews, A. 1996. Semantic case-stacking and inside-out unification. *Australian Journal of Linguistics* 16(1):1–55.
- Austin, P. 1995. Double Case Marking in Kanyara and Mantharta Languages, Western Australia. In F. Plank (Ed.), *Double Case: Agreement by Suffixaufnahme*, 363–79. Oxford: OUP.
- Austin, P., and J. Bresnan. 1996. Non-configurationality in Australian Aboriginal languages. *Natural Language and Linguistic Theory* 14(2):215–268.
- Barron, J. 1998. ‘have’ contraction: explaining ‘trace effects’ in a theory without movement. *Linguistics* 36(2):223–251.
- Borjars, K., N. Vincent, and C. Chapman. 1997. Paradigms, Periphrasis and Pronominal Inflection: a feature-based account. In G. Booij (Ed.), *Yearbook of Morphology 1996*, 155–180. Cambridge, Mass: MIT Press.
- Bresnan, J. 2001. *Lexical Functional Syntax*. Oxford: Blackwell.
- Butt, M., M.-E. Nino, and F. Segond. 1996. Multilingual Processing of Auxiliaries in LFG. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology*, 111–122. Mouton de Gruyter.
- Butt, M. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, CA: CSLI.
- Butt, M., T. H. King, M.-E. Nino, and F. Segond. 1999. *A grammar writer’s cookbook*. Stanford, CA: CSLI Publications.
- Culy, C. 1990. *The syntax and semantics of internally-headed relative clauses*. PhD thesis, Stanford University, Stanford, CA.

- Dalrymple, M. 1993. *The Syntax of Anaphoric Binding*. Stanford, CA: CSLI Publications.
- Dalrymple, M. 2001. *Lexical Functional Grammar*. San Diego: Academic Press.
- Dench, A. 1995a. *Martuthunira: a language of the Pilbara Region of Western Australia*. Canberra: Pacific Linguistics.
- Dench, A. 1995b. Suffixaufnahme and Apparent Ellipsis in Martuthunira. In Frans Plank (Ed.), *Double Case: Agreement by Suffixaufnahme*, 380–95. Oxford: OUP.
- Dench, A., and N. Evans. 1988. Multiple Case-Marking in Australian Languages. *Australian Journal of Linguistics* 8:1–47.
- Evans, N. 1995a. *A Grammar of Kayardild: With Historical-Comparative Notes on Tangkic*. Berlin: Mouton de Gruyter.
- Evans, N. 1995b. Multiple Case in Kayardild: Anti-iconic Suffix ordering and the Diachronic Filter. In Frans Plank (Ed.), *Double Case: Agreement by Suffixaufnahme*, 396–428. Oxford: OUP.
- Frank, A. 2003. Projecting LFG F-structures from Chunks. In M. Butt and T. H. King (Eds.), *Proceedings of LFG 2003*, Stanford, CA. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- Frank, A., and A. Zaenen. 2002. Tense in LFG: Syntax and Morphology. In H. Kamp and U. Reyle (Eds.), *How we say WHEN it happens*, 17–53. Niemeyer.
- Halvorsen, P.-K., and R. M. Kaplan. 1988. Projections and Semantic Description in Lexical-Functional Grammar. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 1116–1122, Tokyo.
- Kaplan, R., and M. Butt. 2002. The Morphology-Syntax Interface in LFG. unpublished handout, LFG 2002.
- Kaplan, R., and P. Newman. 1997. Lexical Resource Reconciliation in the Xerox Linguistic Environment. In D. Estival, A. Lavelli, K. Netter, and F. Pianesi (Eds.), *Proceedings of the ACL Workshop on Computational Environments for Grammar Development and Linguistic Engineering*, 54–61, Madrid.
- King, T. H. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI Publications.
- Lee, H. 1999. The domain of grammatical case in Lexical Functional Grammar. In *Proceedings of LFG99*, Stanford, CA. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- Lee, H. 2002. A Lexicalist OT Approach to Case. unpublished, UNC.
- Nordlinger, R. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI.
- Nordlinger, R. 2000. Australian Case Systems: Towards a Constructive Solution. In Miriam Butt and Tracy Holloway King (Ed.), *Argument Realization*, 41–72. Stanford: CSLI.
- Nordlinger, R., and L. Sadler. 2000. Tense as a Nominal Category. In M. Butt and T. H. King (Eds.), *Proceedings of LFG 2000*. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- O'Connor, R. 2002. Clitics and Phrasal Affixation in Constructive Morphology. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG02 Conference*, Stanford, CA. CSLI Publications: <http://www-csli.stanford.edu/publications>.



- Ørnsnes, B. 2002. Case marking and subject extraction in Danish. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG02 Conference*, Stanford, CA. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- Sadler, L. 1997. Clitics and the structure-function mapping. In M. Butt and T. H. King (Eds.), *Proceedings of LFG97*, Stanford. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- Sadler, L. 1998. English Auxiliaries as Tense Inflections. *Essex Research Reports in Linguistics* 24:1–16.
- Sadler, L., and R. Nordlinger. Under review. Case Stacking in Realizational Morphology. University of Melbourne and the University of Essex.
- Sadler, L., and A. Spencer. 2001. Syntax as an exponent of morphological features. *Yearbook of Morphology 2000* 71–97.
- Sells, P. 2000. Postposing in Japanese. In *Selected Papers from AILA 1999*, 427–442, Tokyo. Waseda University Press.
- Sells, P. to appear. Syntactic Information and its Morphological Expression. In Louisa Sadler and Andrew Spencer (Ed.), *Projecting Morphology*. Stanford, CA: CSLI Publications.
- Sharma, D. 1999. Nominal clitics and constructive morphology in Hindi. In *Proceedings of LFG99*, Stanford, CA. CSLI Publications: <http://www-csli.stanford.edu/publications>.
- Simpson, J. 1983. *Aspects of Warlpiri Morphology and Syntax*. PhD thesis, MIT, Boston, MASS.
- Simpson, J. 1991. *Warlpiri Morpho-Syntax*. Dordrecht: Kluwer Academic Publishers.
- Spencer, A. to appear. Paradigm-based Lexicalism: Negation in Japanese. In Farrell Ackerman and Jim Blevins and Greg Stump (Ed.), *Paradigms and Periphrases*. Stanford, CA: CSLI Publications.
- Stump, G. 2001. *Inflectional Morphology*. Cambridge: CUP.