F-structural Spellout in LFG Morphology

Avery D Andrews
ANU, Mar 2005

Although its architecture ascribes a central role to the lexicon and morphology, most work in LFG has proceeded without reference to any particular theory of morphology. Sadler and Nordlinger (2004, to appear) begin to amend this deficiency by applying Stump's (2001) theory of Paradigm Function Morphology (PFM) to an LFG analysis of some particularly challenging data, the phenomenon of 'case stacking' (Dench and Evans 1988) in Australian languages, especially Martuthunira (Dench 1995) and Kayardild (Evans 1995).[1] In this note I will point out a rather implausible feature of Sadler and Nordlinger's (2004) proposal for connecting the morphological descriptions of PFM to the functional descriptions of LFG, and propose an alternative which I think is simpler and more consistent with what I take to be the basic ideas of PFM.

1 Case-Stacking

The problem of case-stacking is that in languages that have it, inflectional morphology appears to occur in multiple layers, sometimes interleaved with number marking, showing that the inflected word seems to have access to structure several layers above it. A standard entry-level example is this sentence from Martuthunira:

(1)

ngayu	nhawu-lha	ngurnu	tharnta-a	mirtily-marta-a
I	see-PAST	that.ACC	euro-ACC	little-PROP-ACC

thara-nga-marta-a
pouch-LOC-PROP-ACC
I saw the euro with a joey in its pouch
(Dench 1995)

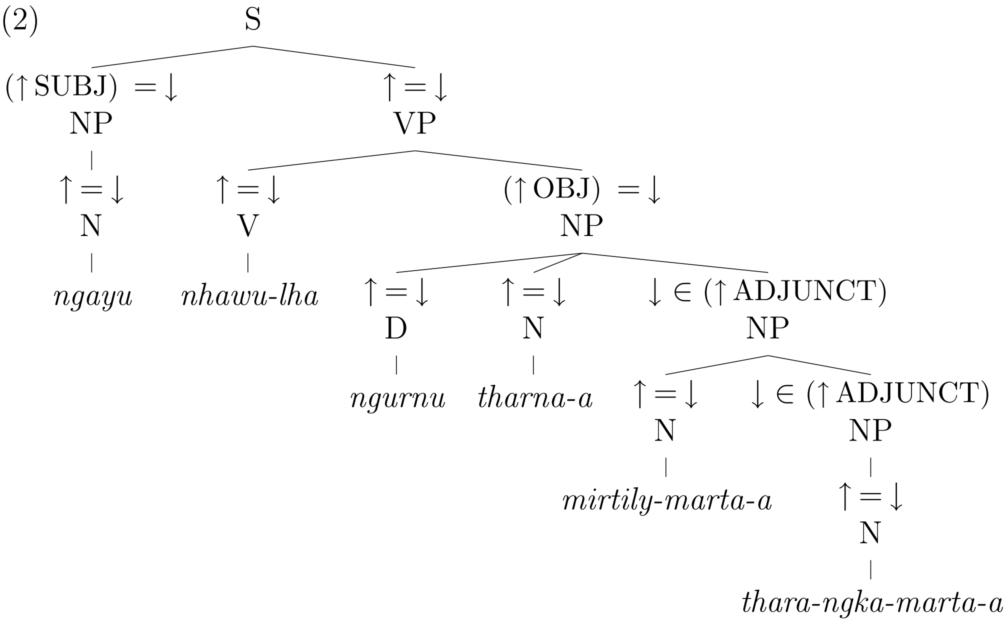
The 'access to higher levels' is indicated by the final two nominals, which constitute a proprietive modifier specifying something that the euro (medium-sized Australian macropod, not European currency) has; the head of this nominal bears first the proprietive case and then an accusative, indicating the function of the NP it is inside of, but this proprietive nominal has a

[1] For more work integrating LFG and PFM, see also Ackerman and Stump (2004).

further locative nominal within it specifying where the little one is. After its locative case-marker, this nominal recapitulates the proprietive and accusative cases.

It seems clear that Martuthunira has NPs, so that the annotated phrase-structure of this example is something like:

(2)



which needs to produce an f-structure like:[2]

[2]I'm assuming that glue-logic or some other kind of semantic interpretation can interpret the semantic cases without requiring Simpson's (1991) idea of having them introduce PRED-features, which complexifies the structures to a considerable extent. I will not however make any specific proposal about this here.

(3)
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Pro'} \\ \text{PERS} & \text{I} \end{bmatrix} \\ \text{TENSE} & \text{PAST} \\ \text{PRED} & \text{`See(SUBJ, OBJ)'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`Euro'} \\ \text{CASE} & \text{ACC} \\ \text{ADJUNCTS} & \left\{ \begin{bmatrix} \text{CASE} & \text{PROP} \\ \text{PRED} & \text{`Little'} \\ \text{ADJUNCTS} & \left\{ \begin{bmatrix} \text{CASE} & \text{LOC} \\ \text{PRED} & \text{Pouch} \end{bmatrix} \right\} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

So how are the nominals going to get access to and spell out properties of higher levels of the grammatical structure, as seems to be happening here?

2 A Proposed Solution

In Sadler and Nordlinger's (2004) analysis, henceforth RMS, this happens in two stages. First, the PFM morphology characterizes a set of 'complete' morphological descriptions, which are pairs consisting of a stem and a set of features, which, in case-stacking languages, have a recursive organization. The morphological descriptions are realized as inflected word-forms by the mechanisms of PFM, and converted to LFG functional descriptions (f-descriptions) by a mechanism described in RMS, which we will shortly be criticizing, and for which we will propose a replacement.

In PFM, a morphological description is a pair consisting of a 'root', representing the lexeme being inflected, and a feature-bundle, or property set, representing the inflectional properties being realized by the item. In original PFM, this is organized as a set of attribute-value pairs essentially as defined in GPSG (Gazdar et al. 1985),[3] but RMS propose the innovation (also in Sadler and Nordlinger to appear) of allowing a feature-bundle to also contain another feature-bundle (only one, which can, however, recur-

[3]With the innovation of classificatory subscripts or distinguishers for attributes, as in AGR(su), or Case_{mod}.

sively contain another, and so on). So the morphological descriptions for *tharnta-a*, *mirtily-marta-a* and *thara-ngka-marta-a*in (1) would be:

(4) a. ⊲tharnta, {Case:Acc}⊳

b. ⊲mirtily, {Case:Prop, {Case:Acc}}⊳

c. ⊲thara, {Case:Loc, {Case:Prop, {Case:Acc}}}⊳

We will have a bit to say later about how these things are produced; the basic idea (Stump 2001:41-42) is that there are 'feature cooccurrence restrictions', like those of GPSG, which say what combinations of features are possible, and a feature bundle (property set) is 'complete' if there is nothing that can be added to it without violating some feature cooccurrence rstriction. Because, unlike most syntactic theories, LFG syntax does not necessarily produce fully specified feature environments for lexical items (number, for example, will normally not be imposed by the syntax, while case is on some analyses but not on others), the notion of 'completeness' is particularly important for applying PFM to LFG: the morphology must produce complete morphological descriptions, with all possible features specified, which are then spelled out as overt forms by morphological realization, and also give rise to f-descriptions in the syntax. If we didn't insist on the descriptions being complete, we would need something like the Morphological Blocking Principle of Andrews (1990) to force forms to carry enough morphological marking, and this principle doesn't appear to be well-regarded in PFM (Ackerman and Stump 2004:130).

Completeness and spellout rules are discussed by Sadler and Nordlinger (to appear); the connection to f-structure is the topic of RMS, and is where Sadler and Nordlinger make a rather questionable move. This is to use the morphological description to produce an annotated tree, reminiscent of what is found in most LFG work, based on the idea of incremental, morph-based morphology, where features are contributed by formatives that are assembled into trees. The annotations on the tree are then solved to produce the f-description, as usual.

For example, here is the tree for (4b) above (I have substituted the standard set-valued treatment of adjuncts for the thematic-role-subscripted ADJ treatment they use):

(5)

N

$\uparrow = \downarrow$	$\uparrow = \downarrow$	$(\leftarrow_s \mathrm{GF}) = \downarrow$
Lex	Case	Case
$(\uparrow$PRED$) =$ 'Little'	$(\uparrow$CASE$) =$ LOC	$(\uparrow$CASE$) =$ ACC
Little	(ADJUNCT $\in \uparrow$)	Acc
	Loc	

Other than the idea of building a tree from a Morphological Description, the only theoretical innovation here is the subscripted left-pointing arrow $\leftarrow_s$, which is supposed to designate the f-structure of the immediately preceeding left-sister (the subscript serves as a reminder that it isn't an off-path equation). These annotations are attached by a principle to the effect that they are used if and only the immediately preceeding sister is a Case affix; otherwise $\uparrow = \downarrow$ is used. GF is a variable over grammatical functions, so what the $(\leftarrow_s \mathrm{GF}) = \downarrow$ annotation says is that the f-structure correspondent of the Case node that it is attached to takes the f-structure correspondent of its left-sister as the value of some unspecified grammatical function (which must be taken to include membership in an ADJUNCT set among the possibilities). That left sister, as well as the stem before it, are both connected to the mother by $\uparrow = \downarrow$, so the result is that the f-structure correspondent of the final case-marker contains the f-structure correspondent of the stem and entire word, as the value of some unspecified grammatical function. Furthermore, the annotation (ADJUNCT $\in \uparrow$) associated with the proprietive case-marker imposes the constraint that this case must be found as a value of the grammatical function ADJUNCT $\in$ within a larger structure.

3 A Difficulty and Proposed Solution

The analysis would appear to work, but the involvement of a tree like (5) is an extremely dubious feature. If the morphological theory we're using is PFM, (5) has no role to play as an analysis of the morphological structure of the word it represents, since this is produced by the spellout rules. It can only serve as an abstract intermediary between the Morphological Description of the word, and the f-description of the sentence. But then why does it have left-right linear order? And why do we need such an abstract intermediary in the first place? Justified abstract intermediaries can be motivated by their role in explaining phenomena. Features, for example, as abstract intermediaries between morphology and syntax, account for how a collection of grammatico-semantic functions, such as perhaps those of the dative case in Latin, can be correlated with a range of different forms of expression (the different morphological markings indicating dative case). But it's not clear at all what the motivation for using a tree as intermediary between syntax and PFM morphology might be, and we can see that there is none by producing an alternative in which the tree is absent, and concomitantly, certain arbitrary decisions don't have to be made, such as whether the tree is head-final or head-initial.

The proposal will be to define a function $\mathcal{F}$ which applies to morphological descriptions directly, and produces f-descriptions as output. Aside from the elimination of the tree, a collateral benefit will be the elimination of the rule deploying the $\leftarrow_s$-annotations, and a more cross-linguistically uniform structural relationship between case and number features.

$\mathcal{F}$ will defined as an 'overloaded' function that does different things to different kinds of inputs. For an entire morphological description, the idea will be to combine the results of converting the root to a PRED-feature with the results of processing the features. A phenomenon which we might well want to capture in the morphology is that of what are from a semantic point of view distinct lexemes having identical morphology; for example *go* as a motion verb and a verb of saying, with the same irregular past *went*. We can accomodate this with a nondeterministic lookup function $\mathcal{L}$ that takes as input a root, and returns as value one of its associated PRED-features, together with any associated syntactic constraints, such as perhaps a control equation, or a 'constructive stem' annotation (RMS:168-170), or an inherent gender specification.

Another complexity is that we will find before too long that feature-processing requires a two-place function with a functional designator as one argument, and a feature-set as the other, the role of the designator being to specify what feature-structure the features in the set are to be added to. We'll use a two-argument case of $\mathcal{F}$ for this, so the definition of $\mathcal{F}$ applied to a morphological description (complete root-property-set pair) will be:

(6) $\mathcal{F}(<L, \sigma>) = \mathcal{L}(L) \cup \mathcal{F}(\uparrow, \sigma)$

σ is then processed by successively consuming its contents, starting with the ordinary attribute-value pairs, and finally dealing with the set-member used to represent case-stacking.

For an ordinary atomic-valued pair $F{:}V$, what we want to do is take the union of the result of processing the pair with the result of processing the remainder, which can be formulated like this:

(7) $\mathcal{F}(f, \{F{:}V\} \uplus \sigma) = \mathcal{F}(f, F{:}V) \cup \mathcal{F}(f, \sigma)$

Here $\uplus$ represents (genuinely, not categorically) disjoint union, so that σ must be a set that doesn't contain $F{:}V$. So successive applications of this remove pairs, and we need a case for when there are none left:

(8) $\mathcal{F}(f, \emptyset) = \emptyset$

(7) invokes a further case of $\mathcal{F}$, applying to a designator and atom-valued pair, which Sadler and Nordlinger (2004:25) specify with a lookup table (which, for them, also provides the annotations for the root).

Our lookup function will be a bit more complicated than theirs, because it has the designator parameter (indicating the f-structure that the annotations will be providing information about), which will be represented here by '$*$' in the annotations (of course we could have used '$\uparrow$', but it seems better to me to use '$*$' so as to indicate that the interpretation is different from that normally used for '$\uparrow$'. So some sample entries for typical morphological features would be:

(9)
$$\begin{array}{lll} \text{Case:Acc} & \Rightarrow & \{(*\,\text{CASE}) \;=\; \text{ACC}, (\text{OBJ}\,*)\} \\ \text{Num:Many} & \Rightarrow & \{(*\,\text{NUM}) \;=\; \text{MANY}\} \\ \text{Case:Prop} & \Rightarrow & \{(*\,\text{CASE}) \;=\; \text{PROP}, (\text{ADJUNCT} \in *)\} \end{array}$$

One might hope that this lookup might be replaced with an universal conversion along the lines of:

(10) $F : V \;\;\Rightarrow\;\; (f\,F) \;=\; V$

But this isn't feasible at the moment because the morphological case features are associated with too much additional material, such as for example the 'constructive case' (Nordlinger 1998) specifications for their environments, illustrated in the Case:Acc entry above by the annotation (OBJ $*$), which has the effect of saying that whatever $*$ gets instantiated as must be the OBJ of something. I'll also point out that if we wanted to account for the meanings of features with meaning-constructors, the entries in this lookup table would be a reasonable place to do it, although developing is beyond the scope of this paper.

$*$ and the designator parameter will turn out to play a rather crucial role in the present account of case-stacking, so it would be good to give them some independent motivation, which can be done by considering the treatment of molecular-valued features, where the value is a subsidiary property-set. These are most clearly needed for treating languages where the verb agrees simultaneously with grammatical properties of both the subject and the object, so that we can't just put, say, subject agreement attributes on the S, VP and V, as was attempted in GPSG.

What Stump does (for example on pp 146, 152) is postulates attributes such as AGR(su) and AGR(ob), which take as values sets of attribute-value pairs, such as $\{\text{PERS:I}, \text{NUM:PL}\}$, which are formally just ordinary feature-property sets. The obvious way to deal with these is to convert the attribute into a grammatical function label, and then process its property-set value in the usual manner, with using something that will designate the value of the grammatical function as first argument.

Although Stump doesn't use conventional LFG GF-labels for the attributes, we can provisionally propose that this is a mere notational difference (although such tricky issues as the correct treatment of inverse forms in Alkonquian and 'version' in Georgian (Anderson 1992) bear on this issue), so we can write things like $\{\text{SUBJ} : \{\text{PERS:I}, \text{NUM:PL}\}\}$ rather than using AGR(su), etc. We can process such things straightforwardly by using a local name as value of the GF, and using that as designator parameter for the recursive processing of the embedded property-set:

(11) $\mathcal{F}(f, GF{:}\tau) = \{(f\,GF) \ = \ \%F\} \cup \mathcal{F}(\%F, \tau)$

A distinct local name will be used to instantiate $\%F$ in each application of (11), we'll indicate the instantiated ones with subscripts.

The result of these provisions is that if the feature-set σ is:

(12) {TENSE:PAST, SUBJ:{PERS:III, NUM:SG}}

then, assuming typical f-descriptional interpretatons for PERS:III and NUM:SG, the result of processing will be:

(13) $$\begin{aligned}\mathcal{F}(\uparrow, \sigma) \quad = \quad & \{(\uparrow\text{TENSE}) = \text{PAST}, (\uparrow\text{SUBJ}) = \ \%F_1, \\ & (\%F_1\,\text{PERS}) \ = \ \text{III}, (\%F_1\,\text{NUM}) \ = \ \text{SG}\}\end{aligned}$$

We have now replicated the functionality of Sadler and Nordlinger's treatment of non-stacked morphology, and extended it to deal with molecular values.

So next we consider the treatment of stacking. Here the morphological property set contains as member a subsidiary property-set, not serving as value of anything. There can be only one of these, and it can recurse structurally (it is certainly an intriguing fact that ordinary outside-in agreement morphology never seems to be recursive; this is only even arguable for 'inside-out' case-stacking). For syntactic realization, we can assume that it is an absolute regularity that the nested property-set isn't dealt with until all of the ordinary attribute-value pairs have been processed (Sadler and Nordlinger deal with iconicity violations in the morphological realization rather than the interface to syntax, and on metatheoretical grounds one would certainly not want the possibility of dealing with it in either place).

The following case of $\mathcal{F}$ will do what needs to be done; it functions in effect as an alternative to (8):

(14) $\mathcal{F}(f, \{\tau\}) = \{\%F = (\mathrm{GF} <\!\in\!> f)\} \cup \mathcal{F}(\%F, \tau)$
(where τ is a property-set)

Applied to (4b), the f-description we'll get is:

(15) $\{(\uparrow\mathrm{PRED}) = \text{`Mirtily'}, (\uparrow\mathrm{CASE}) = \mathrm{PROP}, \%F_1 = (\mathrm{GF} <\!\in\!> f),$
$(\%F_1\,\mathrm{CASE}) = \mathrm{ACC}\}$

This is correct. For (8c), the story would be similar but more complex: there'd be two local names, one set as the inside-out value of the other, each also introducing a case.

Before proceeding to the consideration of number-marking, it is perhaps worth saying a bit about how (15) is supposed to integrate with the rest of the f-description. (15) implements case-stacking by means of 'inside out functional uncertainty', which must apply 'nonconstructively'. This means that an equation such as $\%F = (\mathrm{GF} <\!\in\!> f)$ can only be processed when the developing solution of the f-description contains an f-structure g such that the equation becomes true when g is substituted for $\%F_1$ (this substitution occurs uniformly to all equations in the f-description). The inside-out paths can thus only match against structure that is independently generated in a constructive manner. This would be accomplished by the annotations of the syntactic tree (2). On the other hand, the ordinary 'outside in' equation $(\%F_1\,\mathrm{CASE}) = \mathrm{ACC}$ applies constructively, and the substitution of some other designator for $\%F_1$ may or may not trigger unification of the CASE-value with others introduced independently.[4]

[4]I'm indebted to John Maxwell for some recent discussions of aspects of nonconstructive

4 Number Marking and Well-formedness

An very important aspect of case-stacking, discussed extensively in RMS, is the way in which it can interleave with number marking. This is important because it provides very strong evidence that case-stacking really involves recursive structure, rather than some kind of splitting of the case-attributes into different types, such as syntactic, semantic adverbial, and semantic adnominal (analyses of this nature are discussed and disposed of in rather tedious detail by Andrews 1996). In order to get number-marking to work with the tree-interface to the syntax, Sadler and Nordlinger appear to be forced to put number marking features in distinct layers in the morphological description, as well as case-features, so that the morphological description for a plural accusative nominal would be:

(16) {Num:Pl, {Case:Acc}}

This isn't impossible, but it lacks independent motivation, and would be undesireable for our present treatment, since we would need a more complex convention than (14) to control when iofu annotations are introduced into the f-description, presumably a counterpart of annotation convention (31) of RMS:175.

Fortunately, there's no problem in altering the anaysis so that number and case will be introduced in a flat structure, so that (16) will be replaced with {Num:Pl, Case:Acc}. What we need to do is alter the definition of well-formed morphological property set so that (appropriate kinds of) Case values can also be accompanied by Num values (and therefore, by the idea of completeness, must be so accompanied). RMS doesn't provide an accout of completeness for morphological property sets, and Sadler and Nordlinger (to appear) provides one only for Kayardild, not Martuthunira. In addition to standard 'core' uses of case (syntactic, adverbial, adnominal), Kayardild uses case-marking to express various properties naturally associated with the tense-mood and complementizer systems, which Sadler and Nordlinger distinguish by subscripting the Case attributes as Case_{core}, Case_{mod} (modal case, expressing tense-mood properties), and Case_{assoc} and Case_{comp} (associating and complementizing case, expressing complementizer-system properties). The basic fact is that only Case_{core} can co-occur (at the same level)

annotations.

with a number value, which we can implement by means of the following modification of clause (iii) of (45) in Sadler and Nordlinger (to appear):[5]

(17) If τ is a well-formed property set such that $\mathrm{Case}_{core}{:}\beta \in \tau$, then $\tau \cup \{\mathrm{NUM}{:}\gamma\}$ is a well-formed property set.

In Martuthunira, all cases are core, we will want to allow any property-set specifying Case to extend to one specifying Num.

A property-set that specifies a core case will thus always be extendible to one that specifies a number, so in order to be complete it will have to be so-extended. For the notion of completeness to work, we'll also have to slightly extend the notion of 'extension' on Stump (2001:41) to deal with the recursively embedded property sets, so that if F' is an extension of F, and G' is identical to G except in having F' rather than F as a member, then G' is an extension of G.

Complete property sets will thus always include a number specification as 'sister' to each core case, which will obviously produce an appropriate f-description by our rules. For example the Kayardild nominals:

(18) a. maku-wala-nurru
woman-many-having
having many wives

b. maku-nurru-walad-a
woman-having-many-NOM
many having wives

will arise from the morphological descriptions:[6]

(19) a. <maku, {NUM:MANY, CASE:PROP, {NUM:NEUT, CASE:NOM}}>

b. <maku, {NUM:NEUT, CASE:PROP, {NUM:PL, CASE:NOM}}>

It is clear that these representations will give rise to appropriate f-descriptions.[7]

[5] I've omitted the cardinality parameter because it's unclear to me what its function is, so I'm unsure if it should be incremented (if its purpose is to count feature-value pair tokens) or left the same (if its purpose is to count levels).

[6] I'm here assuming a 'neutral' number when the number value isn't marked (see also the discussion of optionality of numbermarking in Sadler and Nordlinger (to appear:21), fn. 21), and that nominative case can be present but not expressed after an adnominal case-marker.

[7] However, the theory of constructive case and PFM would appear to have some difficulty in accounting for the characteristics of the nominative as an 'elsewhere' case (Evans

5 Conclusion

The conclusion is that the tree-based interface to LFG proposed in RMS can be replaced by a system of what might be called 'functional realization rules', which convert morphological descriptions to functional descriptions by a process similar in spirit to morphological realization, which is mostly universal, but with 'parameters' in the form of specifications of the significance in terms of f-descriptions of particular morphological features. This is simpler than RMS' tree-based approach, and doesn't require arbitrary decisions about linear order. It would seem therefore to be a better idea.

Bibliography

Ackerman, F., and G. Stump. 2004. Paradigms and periphrastic expression. In L. Sadler and A. Spencer (Eds.), 159–185.

Anderson, S. R. 1992. *Amorphous Morphology.* Cambridge University Press.

Andrews, A. D. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 507–557.

Andrews, A. D. 1996. Semantic case-stacking and inside-out unification. *Australian Journal of Linguistics* 1–55.

Dench, A. 1995. *Martuthunira: a Language of the Western Pilbara.* Canberra: Pacific Linguistics.

Dench, A., and N. Evans. 1988. Multiple case-marking in australian languages. *Australian Journal of Linguistics* 1–47.

Evans, N. 1995. *Kayardild: The Language of the Bentick Islanders, with Elements of a Grammar of Proto-Tangic.* Mouton.

1995:136). The syntactic environment for the nominative is (a) subject when there is no complementizing case (b) object when there is neither modal, associative, nor complementizaing case). One can obviously cook up a disjunctive specification to capture this, but this would miss the generalization. If the nominative was a morphological case which introduced no f-descriptional annotations, the Morphological Blocking Principle of Andrews (1990) would be able to capture its distribution, but combining the MBP with PFM would presumably be a last resort, so for now it would be best to wait and see if some other solution will turn up.

Gazdar, G., E. Klein, G. Pullum, and I. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge MA: Harvard University Press.

Nordlinger, R. 1998. *Constructive Case*. Stanford CA: CSLI Publications.

Sadler, L., and R. Nordlinger. 2004. Relating morphology to syntax. In L. Sadler and A. Spencer (Eds.), 159–185. available March 2005 at: `http://www.linguistics.unimelb.edu.au/people/staff/rachel/interfinal.pdf`.

Sadler, L., and R. Nordlinger. to appear. Case-stacking in realizational morphology. to appear in *Linguistics*; prepublication draft available March 2005 at `http://www.linguistics.unimelb.edu.au/people/staff/rachel/PMC-July04.pdf`.

Sadler, L., and A. Spencer. 2004. *Projecting Morphology*. Stanford CA: CSLI Publications.

Simpson, J. H. 1991. *Warlpiri Morpho-Syntax*. Kluwer Academic.

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