

Radical Non-Configurationality without Shuffle Operators

Emily M. Bender, University of Washington

1 Introduction

The word order facts of radically non-configurational languages (including the Australian languages Wambaya [wmb] and Warlpiri [wbp]) pose a challenge to HPSG approaches which assume both that the surface order of words is the yield of the (tectogrammatical) tree and standard HPSG-style cancellation of valence lists. These languages allow discontinuous noun phrases, in which modifiers appear separated from their head nouns by arbitrarily many other words from the same clause. Donohue and Sag (1999) present an analysis based on linearization theory (Reape 1994), which posits that the surface order of words need not be directly determined by the yield of the tree. In this paper, I explore the other alternative: preserving tectogrammatical-phenogrammatical equivalence, and instead accounting for the word order facts of Wambaya with an analysis based on non-cancellation. The analysis described here has been implemented in a medium-sized grammar fragment for Wambaya built on the basis of the LinGO Grammar Matrix (Bender et al. 2002, Bender and Flickinger 2005).

2 Wambaya

Wambaya is a recently-extinct language of the West Barkly family from the Northern Territory in Australia (Nordlinger 1998b). Aside from the constraint that verb-headed clauses require an auxiliary in second position, the word order is free, to the point that noun phrases can be non-contiguous, with head nouns and their modifiers separated by unrelated words. Furthermore, head nouns are generally not required: argument positions that are cross-referenced on the auxiliary can be instantiated by modifiers only, or (for some arguments), if the referent is clear from the context, by no nominal constituent of any kind. There is a rich system of case marking, and adnominal modifiers agree with the heads they modify in case, number, and four genders. An example is given in (1) (Nordlinger 1998b:223).¹ In (1), *ngaragana-nguja* ('grog-propriative', or 'having grog') is a modifier of *ngabulu* milk. They agree in case (accusative) and gender (class IV), but they are not contiguous within the sentence. As with Warlpiri, the pre-auxiliary po-

sition can contain more than one word just in case those words form an NP constituent.

- (1) Ngaragana-nguja ngiy-a
grog-PROP.IV.ACC 3.SG.NM.A-PST
gujunganjanga-ni jiyawu ngabulu.
mother.II.ERG give milk.IV.ACC
'(His) mother gave (him) milk with grog in it.'

3 Previous Analyses

3.1 LFG: Constructive Case

Nordlinger (1998a) presents an analysis of non-configurationality in terms of multiple strategies for the marking of grammatical functions: Configurational languages mark grammatical functions through specific phrase structure positions, while non-configurational languages mark grammatical functions through morphology. Morphological marking of grammatical functions can be through affixes on the verb (head-marking) or on the NPs (dependent-marking).

Nordlinger notes an asymmetry in previous accounts whereby verbal affixes are believed to directly satisfy valence requirements but case markers only match what is provided in the verb's lexical entry, and proposes that instead the case markers should be treated on a par with other kinds of grammatical-function marking morphology in non-configurational languages. In particular, she proposes the following annotated c-structure rules:

- (2) IP → XP I'
 (↑ DF) = ↓ ↑ = ↓
I' → I S
 ↑ = ↓ ↑ = ↓
S → C⁺
 (↑(GF)) = ↓

The I position is filled by the auxiliary. The single position to the left of the auxiliary is filled by a maximal projection assigned some discourse function. The complement of the auxiliary is an endocentric S, consisting of at least one constituent. The S and the auxiliary are f-structure co-heads. Each sub-constituent of S either bears a grammatical function with respect to the clause or is itself the head of the clause. Though this is not explicitly stated in

¹Note that the recipient argument and not the theme is cross-referenced on the auxiliary. However, the third person object marker is a zero. Nonetheless, since first and second person recipients are registered on the auxiliary, the lack of an overt object marker indicates agreement with a third person recipient.

Nordlinger 1998a, the main predicate must also provide a list of grammatical functions, either directly in its lexical entry or indirectly through its a-structure and the general linking theory. The general principles of coherence and completeness require the verb (if there is one) to fill the head role and the nouns to fill argument roles. When there is no verb, a noun can be a predicator.²

The case markers have lexical specifications such as (3) (for case markers on nominal heads) and (4) (for case markers on nominal modifiers):

$$(3) \quad (\text{SUBJ } \uparrow); (\uparrow \text{ CASE}) = \text{ERG}$$

$$(4) \quad (\text{SUBJ } (\text{ADJ } \uparrow)); ((\text{ADJ } \uparrow) \text{ CASE}) = \text{ERG}$$

These inside-out equations allow the case marked nouns and nominals to assert the existence of and associate themselves to an appropriate grammatical function in the clause they belong to (with the modifiers in fact taking an ADJ role inside that grammatical function). Since each nominal does this independently, modifiers and head nouns do not need to be contiguous in the c-structure for their information to be unified at f-structure.

3.2 HPSG: Linearization

Donohue and Sag’s (1999) analysis of Warlpiri would also be applicable to Wambaya. Their analysis based on the DOM feature introduced by Reape (1994). Here, the DOM value of a constituent is a list of *signs*, which include the phonological representations of the words contained within the constituent. Constructions are then classified as either compacting constructions or liberating constructions. Compacting constructions fuse the DOM values of their daughters into a DOM list with a single element. Liberating constructions append the DOM values of their daughters to create multi-element lists. Both types of constructions allow the phonology to be “shuffled”, but only liberating constructions allow their daughters’ phonology to interleave with the phonology from other constituents higher up in the tree.

On Donohue and Sag’s analysis, the NP construction (5) is liberating. That is, it combines two daughters with matching CASE values, but doesn’t constrain them to appear contiguously in the final phonological representation.³ In contrast, the clausal construction (6), which realizes all valence requirements of the head, is a compacting construction.

(5) *mod-nom-cx*:

$$\left[\begin{array}{l} \text{MOTHER} \quad \left[\begin{array}{l} \text{DOM } \delta_1 \circ \delta_2 \\ \text{SYN NP} \end{array} \right] \\ \text{DTRS} \quad \left\langle \left[\begin{array}{l} \text{DOM } \delta_1 \\ \text{SYN } [\text{NP, CASE } \underline{1}] \end{array} \right], \left[\begin{array}{l} \text{DOM } \delta_2 \\ \text{SYN } [\text{NP, CASE } \underline{1}] \end{array} \right] \right\rangle \end{array} \right]$$

(6) *cl-cx*

$$\left[\begin{array}{l} \text{MOTHER} \quad \left[\begin{array}{l} \text{DOM} \quad \left\langle \left[\begin{array}{l} \text{SYN } \underline{0} \\ \text{DOM } \delta_1 \circ \dots \circ \delta_n \end{array} \right] \right\rangle \\ \text{SYN} \quad \underline{0} [\text{VAL } \langle \rangle] \end{array} \right] \\ \text{DTRS} \quad \left\langle \left[\begin{array}{l} \text{DOM} \quad \delta_1 \\ \text{SYN} \quad \left[\begin{array}{l} \textit{finite} \\ \text{VAL} \quad \langle \underline{1}, \dots, \underline{n} \rangle \end{array} \right] \end{array} \right], \dots, \left[\begin{array}{l} \underline{1} [\text{DOM } \delta_2], \dots, \underline{n} [\text{DOM } \delta_n] \end{array} \right] \right\rangle \end{array} \right]$$

This construction is subject to two linear precedence constraints which force auxiliaries to appear before all non-focused elements and a single focused element to precede everything else. These constraints thus determine the auxiliary-second order.

4 Non-Cancellation Analysis

The alternative explored here is based on non-cancellation of valence features. The central intuition of this analysis is that the argument positions of a head are still available for further combination, even after they have been filled and/or subject to modification. This is similar in spirit to Nordlinger’s approach in that it allows the head to combine with its arguments in any order (subject to the auxiliary-second constraint), relying on the matching of case requirements and case marking to sort out which dependent goes with which argument position. This is achieved through altering the head-nexus rules to preserve the SUBJ and COMPS values, and positing new rules which allow modifiers to attach semantically to arguments of the syntactic constituents they combine with. This is reminiscent of work that proposes preserving ARG-ST on phrases (e.g., Frank 1994, Przepiórkowski 2001). However, it differs in using the valence lists themselves to preserve information about a head’s dependents even after they have been

²Though it is not clear how this account captures the fact that nominal predicates don’t co-occur with the auxiliary.

³ \circ represents the operation of ‘domain union’, which shuffles two lists.

realized. Thus this analysis claims that the non-configurationality licensed by non-cancellation correlates with the lack of any notion of saturated or partially saturated constituents.

4.1 Head-arg and head-arg-mod rules

The core of the analysis is thus two series of rules, one for argument realization and one for argument modification. I illustrate here with the rule which targets the second complement position, though there are parallel rules for subjects and the other complement positions. Generalizations across these rules are captured in the type hierarchy. The head-2nd-complement rule is shown in (7). It identifies the SYNSEM value of the non-head daughter with the SYNSEM of the second complement of the head. In addition, it records the information that this argument has been instantiated by its head (rather than just a modifier; [INST +]), and that it has also been instantiated by something ([OPT +]). At the same time, it checks that the argument has not previously been instantiated by its head, by checking for [INST -] on the head-daughter's COMPS list in this position.

(7) *head-2nd-comp-phrase:*

$$\left[\begin{array}{l} \text{SYNSEM} \mid \text{COMPS} \left\langle \boxed{1}, \begin{array}{l} \text{OPT} \quad + \\ \text{INST} \quad + \\ \text{LOCAL} \quad \boxed{3} \\ \text{NON-LOC} \quad \boxed{4} \end{array} \right\rangle \oplus \boxed{A} \\ \text{HD-DTR} \mid \text{COMPS} \left\langle \boxed{1}, \boxed{2}, \begin{array}{l} \text{INST} \quad - \\ \text{LOCAL} \quad \boxed{3} \\ \text{NON-LOC} \quad \boxed{4} \end{array} \right\rangle \oplus \boxed{A} \\ \text{NON-HD-DTR} \mid \text{SYNSEM} \boxed{2} \end{array} \right]$$

The rule which attaches modifiers to the second complement of a head is shown in (8). Like the rule in (7) above, this rule targets the second item on the COMPS list, and provides the information on the mother that it has been overtly realized ([OPT +]).⁴ However, since the non-head daughter is a modifier in this case, rather than matching the non-head daughter's SYNSEM to the complement requirement, it uses the non-head daughter's MOD value instead. This has the effect of giving the modifier the information it

⁴The actual implemented analysis is a bit more complicated than this, since some modifiers are content to appear without the corresponding head in the sentence and others aren't. Space considerations preclude providing full details here.

⁵The specification [FORM *fin*] distinguishes verbs with appropriate inflection for matrix clauses from those inflected to head subordinate modifying clauses. It also has the effect of disallowing auxiliaries as the complement of other auxiliaries, as the auxiliaries are only assigned other values of FORM.

⁶*na* stands for not-applicable. It contrasts with *bool*, which has subtypes + and -. This three-way context is used to similar effect in the English Resource Grammar (Flickinger 2000).

needs about the argument's syntactic and semantic features, to allow agreement in case, number and gender on the one hand, and composition of appropriate semantic representations on the other.

(8) *head-2nd-comp-mod-phrase:*

$$\left[\begin{array}{l} \text{SYNSEM} \mid \text{COMPS} \left\langle \boxed{1}, \begin{array}{l} \text{OPT} \quad + \\ \text{INST} \quad \boxed{5} \\ \text{LOCAL} \quad \boxed{3} \\ \text{NON-LOC} \quad \boxed{4} \end{array} \right\rangle \oplus \boxed{A} \\ \text{HD-DTR} \mid \text{COMPS} \left\langle \boxed{1}, \boxed{2}, \begin{array}{l} \text{INST} \quad \boxed{5} \\ \text{LOCAL} \quad \boxed{3} \\ \text{NON-LOC} \quad \boxed{4} \end{array} \right\rangle \oplus \boxed{A} \\ \text{NON-HD-DTR} \mid \text{MOD} \left\langle \text{LOCAL} \quad \boxed{3} \right\rangle \end{array} \right]$$

4.2 Auxiliaries and word order

Because of the auxiliary-second word order constraint, and because the auxiliaries show agreement with both subjects and objects as well as registering reflexivization, the grammar adopts an argument composition analysis of Wambaya auxiliaries. The auxiliaries all inherit from the following type:⁵

(9) *arg-comp-aux:*

$$\left[\begin{array}{l} \text{HEAD} \left[\textit{verb}, \text{AUX} \quad + \right] \\ \text{SUBJ} \quad \langle \boxed{1} \rangle \\ \text{VAL} \left[\begin{array}{l} \text{COMPS} \left\langle \boxed{2}, \begin{array}{l} \text{OPT} \quad - \\ \text{HEAD} \quad \left[\begin{array}{l} \textit{verb} \\ \text{FROM} \quad \textit{fin} \end{array} \right] \\ \text{SUBJ} \quad \langle \boxed{1} \rangle \\ \text{COMPS} \quad \boxed{A} \end{array} \right\rangle \oplus \boxed{A} \end{array} \right] \end{array} \right]$$

The head-argument and head-arg-modifier phrases illustrated above are all instantiated in both head-final and head-initial versions. The general head-final and head-initial types bear the constraints in (10). These two types use the feature MC ('main clause') to force all constituents to the right of the auxiliary to attach before any to the left, and furthermore, to allow exactly one constituent to attach to the left. That is, an auxiliary plus any number of dependents to the right is [MC *na*],⁶ and a suitable daughter for either

another head-initial rule or the head-final rule. An auxiliary (or auxiliary-headed constituent) that has picked up one dependent to the left is now suitable as either a matrix or a subordinate clause ([*MC bool*]), but can no longer pick up any dependents, since it is now incompatible with the head-daughter position in either head-initial or head-final rules.

$$(10) \quad \begin{array}{l} \textit{aux-head-init:} \\ \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \textit{verb} \\ \text{AUX } + \end{array} \right] \\ \text{MC} \quad \boxed{1} \textit{na} \end{array} \right] \\ \text{HD-DTR } \boxed{2} [\text{CAT} \mid \text{MC } \boxed{1}] \\ \text{NON-HD-DTR } \boxed{3} \\ \text{ARGS } \langle \boxed{2}, \boxed{3} \rangle \end{array} \right] \end{array} \quad \begin{array}{l} \textit{aux-head-final:} \\ \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \textit{verb} \\ \text{AUX } + \end{array} \right] \\ \text{MC} \quad \textit{bool} \end{array} \right] \\ \text{HD-DTR } \boxed{2} [\text{CAT} \mid \text{MC } \textit{na}] \\ \text{NON-HD-DTR } \boxed{3} \\ \text{ARGS } \langle \boxed{3}, \boxed{2} \rangle \end{array} \right] \end{array}$$

4.3 Representations

These aspects of the analysis are implemented together with analyses of a wide-range of phenomena in Wambaya, including argument optionality; subject and object agreement on the auxiliary; various case frames; case, gender and number agreement between nouns and their modifiers; nouns functioning as adverbial modifiers; verbless clauses; coordination; and others. The grammar has been developed against a test suite comprising all of the example sentences in Nordlinger 1998b (800 examples), and presently produces appropriate semantic representations for 91% of these examples, while maintaining relatively low ambiguity (on average 11.89 analyses per item).

In combination with the other analyses in the grammar, the rules and lexical items sketched above assign the structure and the semantic representation in Fig. 1 to the example in (1). The nodes in the tree are labeled with their rule types to better indicate the workings of the analysis.⁷ The most relevant point here is that even though *ngaragananguja* ('with grog in it') and *ngabulu* ('milk') are at opposite ends of the sentence, they both contribute to the semantics of the theme argument (x7 in this example).

5 Comparison

The analysis presented here is, to my knowledge, the first to capture Australian-style discontinuous noun phrases in HPSG without resorting to shuffle or similar operators.⁸ For the core data, it makes the same

predictions as the existing accounts. There are interesting differences, however.

The proposed analysis, like the LFG analysis, allows for NPs to be base-generated as discontinuous. This means that the central problem is linking the pieces back together in the semantics/f-structure. On the present analysis, this is handled by matching constraints on CASE between the verb, the nominal heads, and the modifiers of nominal heads. On Nordlinger's analysis, the verbs have sets of grammatical functions that they require, and the case markers on the nouns/nominal modifiers correlate case to grammatical function. The problem is that the mapping is not one-to-one. As Nordlinger shows, Wambaya has morphological ergativity. This means that both ergative and absolutive case⁹ must be compatible with the grammatical function SUBJ. To avoid getting ergative subjects of intransitives, she has the ergative case marker stipulate the presence of an OBJ function as well. To avoid getting absolutive subjects of transitives, she appeals to 'morphological blocking' (Andrews 1990). She doesn't address semitransitives (with an $\langle \textit{erg}, \textit{dat} \rangle$ case frame), but one possible analysis would be to have dative arguments correspond to an OBJ_θ function rather than plain OBJ. These stipulations are the side-effect of pinning the grammatical function assignment solely on case. It seems to me, however, that the grammatical function requirements of the verbs are a proxy for case, and it would be more straightforward to have the verbs give the case frames directly instead.

Two facts that have to be stipulated in the analysis of Donohue and Sag fall out on this analysis: the fact that the discontinuity is clause-bounded, and the fact that coordinated noun phrases must be contiguous. Donohue and Sag capture these by stipulating the clausal and coordination constructions to be compacting constructions. On the present analysis, discontinuity is clause-bounded because the argument positions are all accessed through the valence lists of the head. Once a particular auxiliary's domain is complete, those arguments are no longer active. As for coordination, since (per typical HPSG assumptions) coordination does not involve modification, there is no way for one coordinand to attach separately from the other. It might be argued that this elegance comes at the cost of having many phrase structure rules which do essentially similar work. I

⁷Some nodes representing lexical rules have been suppressed to simplify the tree structure.

⁸But see Müller 2000 for an account of several related phenomena in German.

⁹Actually, as Nordlinger shows, Wambaya needs a tripartite distinction between ergative, nominative and accusative.

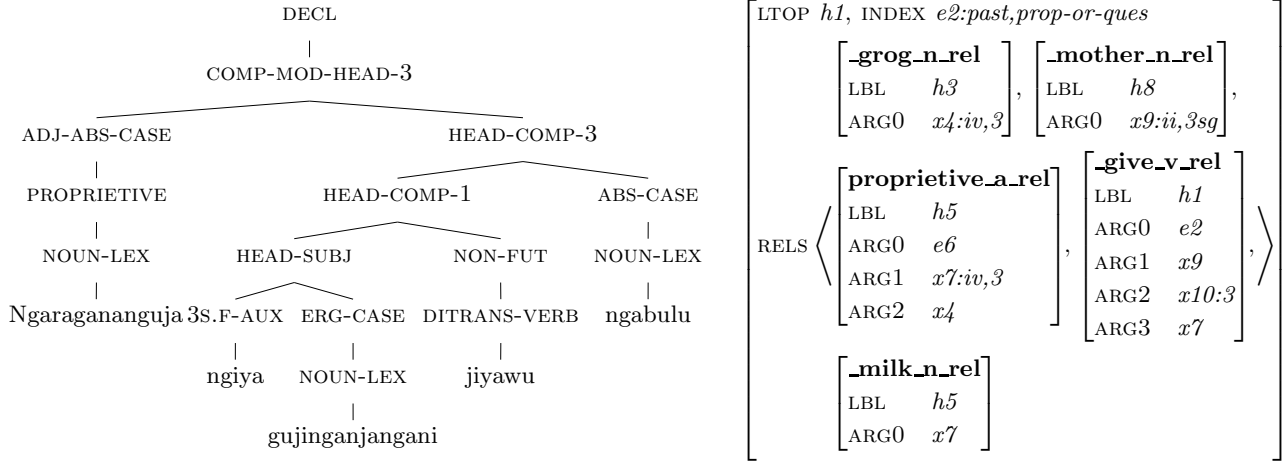


Figure 1: Analysis of example (1)

would reply, however, that series of similar rules are not problematic from an HPSG perspective, so long as the similarities among them are captured in the type hierarchy.

6 Conclusion

The original motivation for approaching this problem was to answer the question of whether Wambaya-style radical non-configurationality could be countenanced within the formalism adopted by the Grammar Matrix (type description language, tdl, as interpreted by the LKB; Copestake 2002). Tdl does not allow for relational constraints; the value of a feature may be identified with the value of another, but not set to some function of the value of the other. Nonetheless, the formalism is Turing complete, and so it is to expected that some analysis is possible. The open question was whether a reasonably elegant analysis was available, and in particular, one which preserves most of the (other) features of HPSG. Even though the analysis here hinges on a radical departure in one sense (non-cancellation of valence arguments), it is reassuring to see that this one departure does not require others. The notions of head-valence rules remain, as do other fundamental notions such as the Head Feature Principle, lexical case assignment and the general mechanisms of semantic compositionality (including the more specific assumptions of Minimal Recursion Semantics (Copestake et al. 2005)). Thus the analysis presented here (and the grammar in which it is implemented) support the idea that the Grammar Matrix and the variety of HPSG it implements can scale to panlingual coverage.

References

- Andrews, A. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8:507–558.
- Bender, E.M., and D. Flickinger. 2005. Rapid prototyping of scalable grammars: Towards modularity in extensions to a language-independent core. In *IJCNLP-05 (Posters/Demos)*.
- Bender, E.M., D. Flickinger, and S. Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Workshop on Grammar Engineering and Evaluation at COLING19*, 8–14.
- Copestake, A. 2002. *Implementing Typed Feature Structure Grammars*. Stanford, CA: CSLI Publications.
- Copestake, A., D. Flickinger, C. Pollard, and I.A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language & Computation* 3(2–3):281–332.
- Donohue, C., and I.A. Sag. 1999. Domains in Warlpiri. Unpublished ms., Paper presented at HPSG’99, Edinburgh.
- Flickinger, D. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6 (1):15–28.
- Frank, A. 1994. Verb second by underspecification. In H. Trost (Ed.), *KONVENS ’94*, 121–130, Berlin. Springer-Verlag.
- Müller, S. 2000. Continuous or discontinuous constituents? In *Proc. ESSLLI-2000 Workshop on Linguistic Theory and Grammar Implementation*, 133–152, Birmingham, UK.
- Nordlinger, R. 1998a. *Constructive Case*. Stanford, CA: CSLI.
- Nordlinger, R. 1998b. *A Grammar of Wambaya, Northern Australia*. Canberra: Research School of Pacific and Asian Studies, The Australian National University.
- Przepiórkowski, A. 2001. ARG-ST on phrases headed by semantically vacuous words: Evidence from polish. In *Proc. HPSG 2000*.
- Reape, M. 1994. Domain union and word order variation in German. In J. Nerbonne, K. Netter, and C. Pollard (Eds.), *German in Head-Driven Phrase Structure Grammar*, 151–197. Stanford: CSLI.