

Complex Predicates in Arrernte

Mark Dras^{1*} François Lareau¹ Benjamin Börschinger¹ Robert Dale¹
Yasaman Motazed¹ Owen Rambow² Myfany Turpin³ Morgan Ulinski²
¹Macquarie University ²Columbia University ³University of Queensland

Using the example of Murrinh-Patha, Seiss [2011] illustrates how Australian Aboriginal languages can shed light on the morphology-syntax interface: one aspect of their polysynthetic nature is that information often encoded in phrases and clauses in other languages is instead found in a single morphological word. In this paper, we look at another instance, the Australian Aboriginal language Arrernte, and in particular at complex predicates within the language, to examine the implications for the morphology-syntax interface. We show how an approach to describing complex verbs in LFG using glue semantics, sketched by Andrews [2007], can be applied to Arrernte complex predicates, and how this fits neatly with the use of glue semantics to model lexical functions in LFG [Lareau et al., 2011].

The context for this work is a data-to-text multilingual natural language generation (MNLG) system, where one of the languages to be generated is Arrernte. (See Wilkins [1989], Henderson [1998] for detailed descriptions and analyses of the language, or Green [1994] for an overview.) For the language realisation component, the grammar is developed in XLE and the morphology in XFST. Some aspects of the language can be handled quite straightforwardly using standard LFG mechanisms. Arrernte at the sentence level is free word order, which we capture by immediate dominance and linear precedence [Falk, 2001]; there's no evidence for constituents like VP, and following [Nordlinger and Bresnan, 2011] we represent the overall c-structure as quite flat. NPs, on the other hand, have relatively fixed internal word order, with the possible exception of some 'floating' of demonstratives and counting terms; we model this again with separation of immediate dominance and linear precedence, but with restrictive linear precedence constraints. NP case (ergative/absolutive, as well as the numerous other cases such as locative, ablative, etc) is handled via 'particles' in the syntax. There is typically no explicit copula: following Nordlinger and Sadler [2007], we use a lexical rule to allow all nominals to act predicatively. For the most part, verb morphology is handled in the morphological component; this includes some relatively complex types of prefixing reduplication, which we handle in XFST using compile-replace rules. In between these clear-cut cases of aspects handled by the grammar versus those handled by the morphology, however, there is the grey area of complex predicates.

Within LFG, a commonly used definition for complex predicates is taken from Butt [1993]. There are three parts to this definition: (a) the argument structure is complex (two or more semantic heads contribute arguments); (b) the grammatical functional structure is that of a simple predicate — it is flat, and there is only a single predicate (i.e. is monoclausal); and (c) the phrase structure may be either simple or complex — it does not necessarily determine the status of the complex predicate. Outside of LFG, there are more general definitions, e.g. by Amberber et al. [2010] in their book on complex predicates, where they acknowledge that there is no agreed set of criteria for defining a complex predicate; Butt [1993] also gives the same caveat. Nordlinger [2010] shows that for the Australian language Wambaya, the associated motion construction under an LFG analysis arguably does not meet Butt's criteria, but the usual classification of associated motion as a complex predicate is still justifiable. In this paper, then, we take the definition of what counts as a complex predicate from Henderson [2002], where he uses phonological, morphological (e.g. the possibility of reduplication) and grammatical criteria to define the boundaries.

Given this definition of complex predicate for Arrernte, we look at three particular types: associated motion in (1), intransitive verbalisers in (2), and transitive verbalisers in (3).

- (1) a. *artwe angk-artn.alpe-ke*
man speak-Quick:DO&GO.BACK-PC
The man quickly spoke and then went back.
- b. *are-ty-arle akwele alh-err-eme*
see-PRIOR.MOTION-FOC SUPPO GO-DUAL-NPP
Two supposedly go and see.
- c. *artwe angk-intye-ke*
man speak-DO.COMING-PC
The man spoke while coming this way.
- (2) a. *artwe akngerre-irre-me*
man big-IV-NP
The man is becoming big.
- b. *alakenhe re ampe akweke mpwe ulk-etyenh-ele*
thus 3sg:NOM child small urine excrete-FUT-SS
irr-entye.akngerre
IV-NOMLSR
Little kids behave that way when they need to have a leak.
- (3) a. *mperlk-elhe anthurre renhe il-eme*
be.white-TV INTENS 3sg:ACC TV-NP
make it go really white

In associated motion constructions, “a verb-stem action happens against the background of a motion event with a specific orientation in space” [Wilkins, 2006]. Wilkins [1989] sees the associated motion as a morpheme for which there is a specific slot in the verb stem. Based on further data, Henderson [2002], on the other hand, notes that the construction e.g. in (1a) could instead be glossed as an associated motion particle *-artn-* and the full verb of motion *alpeme* ‘to go and come back’; that some intervening material is possible, e.g. the focus and supposition particles in (1b); and that there are phonological grounds for considering them separate words. However, the morphological criteria suggest that the construction is a single word, and of a hierarchy of possible intervening material for complex predicates given by Henderson [2002] ranging from most to least restricted, only the most restricted can be used for associated motion. Further, and most relevant for implementation in a computational grammar, a number of types of associated motion constructions, as in (1c), cannot be decomposed into a separate particle and a full verb. Hence, we leave this type of complex predicate to the morphological component.

This contrasts quite strongly with the intransitive verbaliser *irreme*, which often functions as an inchoative, transforming e.g. nominals into intransitive verbs. This kind of complex predicate permits the widest variety of intervening material, as in (2b). In that example, *irreme* is attached to

*Corresponding author: Department of Computing, Macquarie University, NSW 2109 Australia. Mark.Dras@mq.edu.au

alakenhe ‘thus’, with a gloss of ‘behave in that way’; a full dependent clause intervenes. The transitive verbaliser *ileme*, which often functions as a causative, transforming e.g. nominals or intransitive verbs into transitive verbs, is similar, although the amount of intervening material it permits is somewhat smaller. This suggests that these verbalisers could be acting in a manner similar to light verbs, which motivates our analysis of this second broad class of complex predicates.

Complex predicates in LFG have most often been handled using some kind of predicate composition, sometimes at the level of argument structure elaborated via Jackendoff’s Lexical Conceptual Structures (LCS) [Jackendoff, 2002], and in some cases through the use of the restriction operation [Butt, 1993, Alsina, 1997, Andrews and Manning, 1999, Nordlinger, 2010, exemplify some approaches]. An alternative proposed by Andrews [2007] is to use glue semantics, which is outlined there with a sketch of Romance causatives. For our context of an MNLG system, this is appealing. Lareau et al. [2011] note that it has been the experience of large-scale MNLG systems [Wanner et al., 2010, for example] that as much of the system as possible should be language-independent; to this end, they propose the incorporation of so-called lexical functions from Meaning-Text Theory (MTT) [Kahane and Polguère, 2001], which embody recurrent patterns of collocations. These abstract away from language-dependent collocations, such as the English *outright lie* versus French *mensonge éhonté* ‘shameless lie’, as well as language-internal collocational variation, such as *heavy rain*, *strong wind* or *intense bombardment* which all refer to the intensification of some phenomenon. This particular semantic notion of intensification or strength is represented by $\text{Magn}(L)$; another lexical function of interest is $\text{Oper}_1(L)$, where a semantically (mostly) empty verb serves as syntactic support to link a predicative noun to its most prominent semantic argument, for example $\text{Oper}_1(\text{TALK})=\text{GIVE}$, $\text{Oper}_1(\text{ATTENTION})=\text{PAY}$. Using glue semantics to ensure the mapping between semantic arguments and syntactic functions à la Dalrymple [2001], Lareau et al. [2011] define these as lexical rule templates.

We can use this exact same mechanism to handle our second broad class of complex predicates in the grammar component. In some ways, as Andrews [2007] notes (albeit with a different formulation of glue semantics in that paper), glue semantics is mimicking the effect of the alternative LCS approach at a-structure. For Arrernte intransitive verbalisers, the relevant lexical function is $\text{IncepOper}_1(L)$, similar to $\text{Oper}_1(L)$ but referring to a support-like verb indicating the start of something (e.g. *contract a disease*). We define the template INCEPOPER (below left). For sentence (2a) above, *irreme* would instantiate this template; *akngerre* is a nominal that is verbalised by *irreme*. For the generalised f-structure and lexical items with semantics below centre, we obtain the desired overall semantics for the complex predicate (below right).

<p>INCEPOPER1(L)= $(\uparrow \text{PRED}) = \text{‘IncepOper}_1(\uparrow \text{SUBJ}, (\uparrow \text{OBJ}))\text{’}$ $(\uparrow \text{OBJ PRED}) = \text{c‘L’}$ $\lambda X.\text{start}(X) : (\uparrow \text{OBJ})_\sigma \multimap \uparrow_\sigma$</p>	$\left[\begin{array}{l} \text{PRED} \quad \text{‘IncepOper}_1(2:\text{artwe}, 3:\text{akngerre})\text{’} \\ \text{TENSE} \quad \text{pres} \\ \text{SUBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{‘artwe’} \\ \text{PERS} \quad 3 \\ \text{NUM} \quad \text{sg} \end{array} \right] \\ \text{OBJ} \quad \left[\begin{array}{l} \text{PRED} \quad \text{‘akngerre’} \end{array} \right] \end{array} \right]$	<p>[artwe] $\text{man} : \uparrow_\sigma$ [akngerre] $\lambda X.\text{big}(X) : ((\text{OBJ}\uparrow) \text{SUBJ})_\sigma \multimap \uparrow_\sigma$ [irreme] $\lambda X.\text{start}(X) : (\uparrow \text{OBJ})_\sigma \multimap \uparrow_\sigma$</p>
---	---	--

This could be tidied up further. Because *irreme* is very productive, most nominals will have both their regular semantics and the one given above for *akngerre*, suggesting a lexical rule along the lines of the one handling the absent copula. The transitive verbaliser works similarly, and the analysis has a number of similarities with the one proposed for Romance causatives in Andrews [2007]. Overall, then, even though there is a perhaps unexpected division among Arrernte complex predicate types with respect to the grammar and morphology components, the analysis given above harmonises with a role for glue semantics in LFG and a proposed use of LFG in MNLG systems.

References

- Alex Alsina. A Theory of Complex Predicates: Evidence from Causatives in Bantu and Romance. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 203–246. CSLI Publications, Stanford, CA, US, 1997.
- Mengistu Amberber, Brett Baker, and Mark Harvey, editors. *Complex Predicates*. Cambridge University Press, Cambridge, UK, 2010.
- Avery Andrews. Projections and Glue for Clause-Union Complex Predicates. In *Proceedings of the LFG07 Conference*, pages 44–65, 2007.
- Avery Andrews and Christopher Manning. *Complex Predicates and Information Spreading in LFG*. CSLI Publications, Stanford, CA, US, 1999.
- Miriam Butt. *The Structure of Complex Predicates in Urdu*. PhD thesis, Stanford University, 1993.
- Mary Dalrymple. *Lexical Functional Grammar*, volume 42 of *Syntax and Semantics Series*. Academic Press, New York, 2001.
- Yehuda Falk. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. CSLI Publications, Stanford, CA, US, 2001.
- Jenny Green. *A Learner’s Guide to Eastern and Central Arrernte*. IAD Press, Alice Springs, Australia, 1994.
- John Henderson. *Topics in Eastern and Central Arrernte grammar*. PhD thesis, University of Western Australia, 1998.
- John Henderson. The Word in Eastern/Central Arrernte. In R. M. W. Dixon and Alexandra Aikhenvald, editors, *Word: A Cross-Linguistic Typology*, pages 100–124. Cambridge University Press, Cambridge, UK, 2002.
- Ray Jackendoff. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford, UK, 2002.
- Sylvain Kahane and Alain Polguère. Formal foundation of lexical functions. In *Proceedings of ACL 2001*, Toulouse, 2001.
- Francois Lareau, Mark Dras, Benjamin Börschinger, and Robert Dale. Collocations in Multilingual Natural Language Generation: Lexical Functions meet Lexical Functional Grammar. In *Proceedings of the Ninth Australasian Language Technology Workshop (ALTA’11)*, pages 95–104, Canberra, Australia, 2011.
- Rachel Nordlinger. Complex predicates in Wambaya: detaching predicate composition from syntactic structure. In Mengistu Amberber, Brett Baker, and Mark Harvey, editors, *Complex Predicates*, pages 237–258. Cambridge University Press, Cambridge, UK, 2010.
- Rachel Nordlinger and Joan Bresnan. Lexical-Functional Grammar: interactions between morphology and syntax. In Robert Borsley and Kersti Börjars, editors, *Non-Transformational Syntax: Formal and Explicit Models of Grammar*. Wiley-Blackwell, 2011.
- Rachel Nordlinger and Louisa Sadler. Verbless clauses: revealing the structure within. In J. Grimshaw, J. Maling, C. Manning, J. Simpson, and A. Zaenen, editors, *Architectures, Rules and Preferences: A Festschrift for Joan Bresnan*, pages 139–160. CSLI Publications, Stanford, CA, USA, 2007.
- Melanie Seiss. Implementing the Morphology-Syntax Interface: Challenges from Murrinh-Patha Verbs. In *Proceedings of the LFG11 Conference*, pages 482–502, 2011.
- Leo Wanner, Bernd Bohnet, Nadjat Bouayad-Agha, François Lareau, and Daniel Nicklaß. MARQUIS: Generation of user-tailored multilingual air quality bulletins. *Applied Artificial Intelligence*, 24(10):914–952, 2010. doi: 10.1080/08839514.2010.529258.
- David Wilkins. *Mparntwe Arrernte (Aranda): studies in the structure and semantics of grammar*. PhD thesis, Australian National University, 1989.
- David Wilkins. Towards an Arrernte grammar of space. In Stephen Levinson and David Wilkins, editors, *Grammars of space: explorations in cognitive diversity*, pages 24–62. Cambridge University Press, Cambridge, UK, 2006.