Degema and the string interface
Tina Bögel
University of Konstanz

Recent discussion of prosodic analyses with LFG (Bögel et al. 2009, Dalrymple and Mycock 2011 (henceforth DM11)), has generated interest in the function and the nature of the string in its interaction with the lexicon. In general, a syntactically unparsed, but tokenized string is taken to be mapped to c-structure via the relation π (Kaplan 1987, Asudeh 2006), where c-structure encodes the linear order and hierarchical structure of the elements contained in it (Dalrymple 2001). However, exceptions to the strict linearity of the relationship between the string and the c-structure have been proposed. One well-known case is the prosodic inversion that has been proposed to account for second position clitics (e.g., Bögel et al. 2010, following Halpern 1995). Alternatively, lexical sharing has been proposed (Wescoat 2009). This has the advantage of respecting linearity, but the disadvantage of positing additional formal machinery.

The approach in this paper, in contrast, makes use of the multi-dimensional nature of the lexicon within LFG and shows how an explicit acknowledgement of an interaction with postlexical rules together with the availability of inside-out functional uncertainty (IO-FU) can deal with the notoriously difficult problem of endoclisis without the positioning of extra formal machinery or a violation of linearity in mapping from the string to c-structure. In particular, the analysis assumes the architecture proposed by DM11 in which the string is seen as consisting of two parts: the p(honological)-string and the s(yntactic)-string. They are related to one another via a multidimensional lexicon.

The paper shows how to analyze endoclisis in Degema (spoken in Nigeria). Degema is a tone language which distinguishes between some segmentally and categorical identical lexical entries only via lexical tone. As a consequence, a lexical entry must contain information about tone. In Figure 1, the Degema word *ugo* (ambiguous between 'vulture' and 'butterfly') is given as a speech signal, which, in the approach to the string interface presented in this paper, is equal to the p-string. The speech signal is analysed in the p-diagram, a compact representation of the speech signal and an interpretation thereof (Bögel 2012).

lexicon (p-form)	p-string (= speech signal)	p-structure/p-diagram

TONES	L	H
VALUE	/u/	/go/
V-INDEX	S_1	S_2

< ugó> LEXICAL TONE LH ...
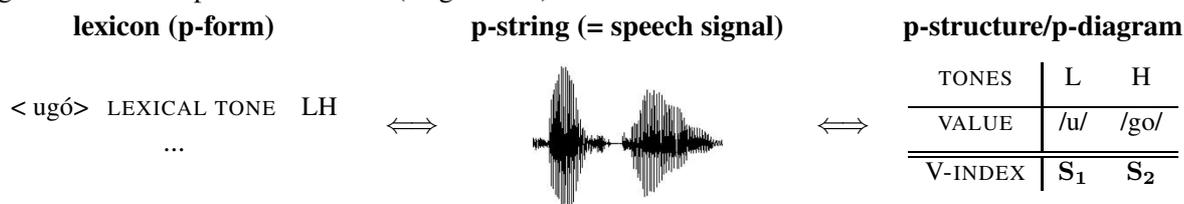
Figure 1: Lexical entry, p-string and p-diagram of the word *ugo* (here: 'vulture').

Within LFG, the lexicon has always been treated as cutting across modules (at least morphology and syntax, but also semantics). Roughly following Levelt et al. (1999), the lexicon presented here is divided into three parts: (1) the conceptual interpretation (semantics), (2) the morphosyntactic information and (3), the phonological aspects of a lexical entry. This paper focuses on the definition of the morphosyntactic and the phonological properties (s-form and p-form, cf. DM11). Tabel 1 shows the s-form of a segmentally identical word (*ugo*), and how the p-form can disambiguate segmentally identical strings by means of lexical tone. (Degema has three tones: High (x́), Low (x), and downstepped High ($^{\downarrow}$x), which is a High tone following another one.)

concept	s-form		p-form		
PALM-NUT VULTURE	ugo N (↑ PRED) = 'ugo'		< ugó>	SEGMENTS METRICAL FRAME LEXICAL TONE	/u g o/ ($\sigma\sigma$) LH
BUTTERFLY	ugo N (↑ PRED) = 'ugo'		< úgo>	SEGMENTS METRICAL FRAME LEXICAL TONE	/u g o/ ($\sigma\sigma$) HL
A KIND OF STEW	ugo N (↑ PRED) = 'ugo'		< úg$^{\downarrow}$o>	SEGMENTS METRICAL FRAME LEXICAL TONE	/u g o/ ($\sigma\sigma$) H$^{\downarrow}$H

Table 1: Lexical entries for *ugó* 'vulture' and *úg$^{\downarrow}$o* 'butterfly'.

The p-form entry has three attribute-value pairs. First, the segments of the string are listed (these should actually be realized as phonetic feature bundles). Second, the metrical frame of the entry states the number of syllables among which the segments are distributed. The brackets around the syllables show that the lexical entry is a prosodically complete 'word', meaning that the lexical entry does not depend on a host (as, e.g., clitics do). Third, there is lexical tone, which, in this case, is the only difference between the three entries for *ugo*.

Now consider Degema endoclitics with this type of multidimensional lexicon in mind. Degema has several types of clitics. One group are enclitics, which occur after their host. Of these, one is the factative clitic (expresses that

a given event/state holds). The factative clitic consists of an underspecified Vowel *V* and *n*: *Vn*. In a clause-final position and after a host ending in a consonant, it appears as an endoclitic within its host. (1) shows an example of the factative enclitic in clause-final position, following a host ending in a vowel. (2) shows the factative clitic in an endoclitic position following a host ending in a consonant.

(1) mi=   bí   =$^{\downarrow}$**in**   (2) ɔ=   nʊ́   ba̩=$^{\downarrow}$**a_w**
     1SgP= be black =fact   3SgP= hit   them$_1$=fact_them$_2$
     'I am black'   (Kari 2002, 43)   'S/he hit them'   (Kari 2002, 45)

In (1) and (2) the underspecified vowel of the clitic takes on the qualities of the host's vowel via vowel harmony. Furthermore, while the host's lexical tone remains 'High', the clitic's underspecified vowel receives a downstepped high tone ($^{\downarrow}$). In (1) the clitic is attached after the host, but appears inside of the host in (2), switching position with the host's final consonant and with the clitic's final /n/ deleted (Kari 2004, 396/397).

This process of endocliticisation can be analysed by means of the multi-dimensional lexicon, an assumption of a postlexical rule component which interacts with it and existing ideas within LFG. To begin with the lexicon, Table 2 shows the lexical entries for the factative clitic and the pronominal host. The lexical entry indicates that the clitic is a prosodically deficient item, because the syllable is not surrounded by brackets. The direction of cliticisation (enclitic) is indicated by the left hyphen.

s-form		p-form		
ba̩w PRN ($\uparrow$ PRED) = 'they'		< bá̩w>	SEGMENTS	/ba̩ a w/
...			METRICAL FRAME	(σ)
...			LEXICAL TONE	H
Vn CL ($\uparrow$ EVENT-TYPE) = factative		< V̌n>	SEGMENTS	/V n/
...			METRICAL FRAME	$-\sigma$
...			LEXICAL TONE	H

Table 2: Lexical entries for the factative clitic and the pronoun *ba̩w* 'they'.

However, this does not as yet explain how the clitic is integrated into the host. This process is taken to be part of a set of postlexical rules that form a phonologically acceptable string. The postlexical rules consider whether: 1) the enclitic is in sentence-final position, 2) the host has a VC ending, an information, which is encoded in the lexicon. Further relevant postlexical rules are the vowel harmony and that two adjacent High tones will cause the second High tone to be downstepped.

The syntactic analysis of the endoclitic phenomena poses a different kind of challenge. Here the problem is that the f-description of the endoclitic describes information about the matrix predicate (the event/state) but that the endoclitic appears within the lexical form of the object (other possibilities are other grammatical functions or the verb itself). However, the difficulty that the information associated with the clitic is potentially "trapped" within an embedded f-structure can be dealt with via the application of IO-FU. Consider Figure 2, where the clitic appears inside of the object pronoun. The clitic contributes information that the EVENT-TYPE is factive, but that this information should pertain to a path which goes via the object to the other f-structure. As the f-structure in Figure 2 shows, this yields exactly the right result.

Figure 2: Endoclisis, from the lexicon to f-structure.

In conclusion, I have formalized a string-lexicon-postlexical component interface. I have shown that this enables the analyses of phenomena like lexical tones and endoclitics, without altering the linear order of the string, but maintaining a parallelity between the s-string and the corresponding p-string, the speech signal.

References:

Asudeh, A. 2006. Direct Compositionality and the Architecture of LFG. In Miriam Butt, Mary Dalrymple, and Tracy Holloway King (eds.), *Intelligent Linguistic Architectures: Variations on Themes by Ronald M. Kaplan*, 363Ð387. Stanford, CA: CSLI Publications.

Bögel, T. 2012. The P-diagram – a Syllable-based Approach to P-structure. *Proceedings of LFG12*.

Bögel, T., M. Butt, R. Kaplan, T.H. King and J. Maxwell III. 2010. Second Position and the Prosody-Syntax Interface. *Proceedings of the LFG10 Conference*, Carleton University, Ottawa.

Bögel, T., M. Butt, R. Kaplan, T.H. King and J. Maxwell III. 2009. Prosodic Phonology in LFG: A New Proposal. *Proceedings of LFG09*.

Dalrymple, M. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Dalrymple, M. and L. Mycock. 2011. The Prosody-Semantics Interface. *Proceedings of LFG2011*.

Halpern, A. 1995. *On the Placement and Morphology of Clitics*. CSLI Publications.

Kaplan, R. M. 1987. Three Seductions of Computational Psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson, and Paul Bennett (eds.), *Linguistic Theory and Computer Applications*, 149Ð181. London: Academic Press.

Kari, E. 2004. *A reference grammar of Degema*. Köln: Köppe.

Kari, E. 2002. On Endoclitics: Some facts from Degema. *Journal of Asian and African Studies*, 63.

Levelt,W., Roelofs, A. and A. Meyer. 1999. A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences* 22, 1Ð75.

Wescoat, M. T. 2009. Udi Person Markers and Lexical Integrity. *Proceedings of LFG09*, CSLI Publications.