

# Individuating lexemes

Andrew Spencer, University of Essex  
LFG15 Meeting, Waseda University, Tokyo

19 June 2015

A perennial problem for the lexicographer is deciding when two closely related words constitute two related forms of a single lexeme/lexical entry or two distinct but related lexical entries. For instance, are the transitive and intransitive uses of *BREAK* two entries or one? In a lexeme-based model of morphology this question takes a variety of forms and raises a variety of (difficult) questions. One particularly interesting question centres around the notion of ‘transposition’. For instance, in a language with distinct adjective and verb morphology, is a deverbal participle a distinct lexeme/lexical entry or is it a form of the base verb lexeme? On the one hand the participle inflects like an adjective not a verb, but on the other hand the participle typically retains verb properties such as argument structure, and even idiosyncratic case marking, as well as verbal inflectional categories such as voice, aspect or even tense. I will argue that this problem is intimately related to the question of what constitutes a lexeme/lexical entry (cf Spencer 2013). A true transposition of category A to category B has to have its own lexical entry (that of B), but it has to be a form of a category A lexeme. This means we have to have (i) a way of defining lexemes/lexical entries and (ii) a way of individuating them.

Minimalism/Distributed Morphology takes a radical stance on individuating lexemes: there are no ‘words’ of any kind, either lexemes or (inflected) forms of lexemes, there are only categoryless roots identified solely by their phonology (Borer, 2013). This means that it would be easy to define a participle as a root which first merges with a *v* head and later with an *a* (adjective) head, but this leads to the (unsolved) problem of overgeneration. In addition, the strict interpretation of the DM account brings the intriguing entailment that there can be no true lexical suppletion of the *go~went* variety (Borer 2013) – only ‘functional elements’ can show suppletion, contrary to easily established fact (Harley 2014, Spencer, to appear).

There seem to be no systematic studies of transpositions in HPSG, apart from treatments of action nominals as mixed category types. The question of defining and individuating lexemes is addressed in Sag’s (2012) model of Sign-Based Construction Grammar (SBCG). A lexeme is a sign, a pairing of a form with a meaning (and other properties). It is therefore a linguistic object. A lexical entry is a description, not an object, and hence not a sign. Rather, it is a type of construction. But a lexeme is supposed to be a type of sign, even though it has no single form and cannot actually correspond to a linguistic object until inflected, at which point it is a word form and not a lexeme. The notions of lexeme, lexemic identity

and lexical entry are thus somewhat confused in SBCG. In effect, there is no conceptual difference between the type 'lexeme' and the type 'lexical entry', contrary to assumption. These inconsistencies make it difficult to see how SBCG could account for transpositions and the difficulties are probably similar in other varieties of HPSG.

Sag deploys a Lexical Identifier (LID) as a 'housekeeping' feature. The LID is a property of the SYN|CAT attribute and it effectively recapitulates the semantic representation of the lexeme. However, there are serious problems with this approach. First there's no good reason to think of lexemic identity as a SYN property (rather, it's a property of the lexical entry/lexeme as a whole). Second, lexemic identity is surely independent of meaning (cf Bonami, 2015): two perfect synonyms can still be distinct lexemes (for instance, belonging to distinct morphological, inflectional classes), but more subtly, we find that there are distinct lexemes which do not differ in meaning but which are nonetheless different lexical entries. A case in point seems to be the denominal relational adjectives of English, Russian and many other European languages, which don't differ in cognitive content from the nouns from which they are derived but which are clearly distinct lexemes: *preposition prepositional*, *tide tidal*, *Sun solar*, *spring vernal* etc, etc. These can be contrasted with the relational adjective transpositions of Selkup, which are actually inflected forms of the noun lexeme (Spencer 2013).

In Spencer (2013) I propose that each lexeme/lexical entry is individuated by means of a Lexemic Index (LI), an integer unique to each lexeme, similar, but not identical, to Stump's (2001) notion of 'lexical index', and Bonami's (2015) LID attribute. The LI is a distinct attribute, independent of phonology, morphology and semantics (though in Spencer, to appear, I argue that entries with a distinct LI should differ at least minimally in semantic interpretation). I briefly summarize the way that this LI can be deployed to capture a variety of types of lexical relatedness.

Traditionally, LFG lexemes have been individuated by their PRED values without the need for a special lexemic index. The PRED attribute plays an important role in excluding structures which would involve a PRED value clash. However, a blanket ban on lexemes with two PRED values poses problems for the analysis of complex predicates (Andrews and Manning, 1999). Moreover, it's far from clear what the relationship should be between the PRED value and the lexical semantic representation, given the factorization of information into morphological, syntactic, semantic etc which is characteristic of LFG. However, most of the traditional functions of the PRED attribute seem to have been distributed to other parts of the architecture: the Glue language accounts for Consistency/Coherence (Andrews, 2007; Asudeh, 2012) and Predicate Uniqueness can be made to follow from the way lexemes are instantiated in syntactic structure. I therefore propose that we re-interpret the PRED attribute as equivalent to the LI attribute in the lexical representations argued for in Spencer (2013). I show how the mechanisms of lexemic individuation and at least some of the principles of lexical organization proposed in Spencer (2013) can therefore be straightforwardly integrated into the LFG architecture. For standard cases of the morphology-syntax interface this has little effect and standard analyses are preserved, but it provides the basis for an explicit treatment for non-standard word types such as transpositions (see Spencer, talk to be given at morphology workshop, LFG15).

## References

- Andrews, A. (2007). Projections and glue for clause-union complex predicates. In Butt, M. and King, T. H., editors, *The Proceedings of the LFG07 Conference*, page 44–65, University of Stanford, California, USA.
- Andrews, A. and Manning, C. (1999). *Complex Predicates and Information Spreading in LFG*. CSLI Publications, Stanford, CA.
- Asudeh, A. (2012). *The Logic of Pronominal Resumption*. Oxford University Press, Oxford.
- Bonami, O. (2015). Periphrasis as collocation. *Morphology*, 25:63–110.
- Borer, H. (2013). *Structuring Sense: Taking Form*. Oxford University Press, Oxford.
- Harley, H. (2014). On the identity of roots. *Theoretical Linguistics*, 40:225–275.
- Sag, I. A. (2012). Sign-Based Construction Grammar: an informal synopsis. In Boas, H. C. and Sag, I. A., editors, *Sign-Based Construction Grammar*, page 69–202. CSLI Publications, Stanford, CA.
- Spencer, A. (2013). *Lexical Relatedness: A Paradigm-based Model*. Oxford University Press, Oxford.
- Spencer, A. (to appear). How to individuate lexemes. In Siddiqi, D. and Harley, H., editors, *Morphological Metatheory*, page xxx–xxx. John Benjamins Publishing Co., Amsterdam.
- Stump, G. T. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge University Press, Cambridge.