

Position paper: 'Grammatical semantics and multilinguality: what stands behind the lexicon?'

John A. Bateman, Project KOMET, GMD/IPSI, Darmstadt, Germany

There has in recent years been a steady increase in the role given to the lexicon in computational linguistics. Accordingly, there are now also many efforts to uncover appropriate organizations of lexical information: including proposals for taxonomies of semantic organizational primitives/features, 'ontology' design, etc. This very necessary activity seems to me, however, to be partly compromised by a second trend also resulting from the attention given to the lexicon. That is the move to lexicalize grammars so that the 'grammatical' component becomes minimal and grammatical properties are 'projected' from those of their lexical components. By reducing the role of grammatical considerations, a strong source of information about useful lexical organization has been removed. Although this is obscured when the kind of grammar that is worked with is structural and constituency-based, many of the efforts to provide lexical organization are attempting to recover the kind of information that functional *clause*-based grammars consider basic. This is because it is the *clause* that provides the most natural link upwards to semantics and context, not the, more arbitrarily related, word.

In functional clause-based grammars and semantics such as those pursued in systemic-functional linguistics (e.g., the English grammar of the Penman project and the German grammar of the KOMET project), linguistic description is broken down according to strata (context, semantics, lexicogrammar, phonology). Each stratum receives a complex description in terms of a subsumption lattice of *functional paradigmatic* alternatives — paradigmatic in that the focus is on linguistic *alternations*, i.e., the kind of relations that are in lexicalized grammars often placed in 'lexical rules', and functional in that the alternations are classified according to contrasts in communicative function. Within a systemic grammar *all* linguistic phenomena are described in terms of alternations — the current grammar of English, for example, contains about 600 'lexicogrammatical' alternations, 600 semantic alternations, and 250 'ontological' alternations. In addition, each stratum is linked both to the strata above and the strata below by means of inter-stratal realization statements. For details and discussion of all these constructs, see Matthiessen and Bateman (1991: *Text generation and systemic linguistics: experiences from English and Japanese*, New York: St. Martin's Press).

The subsumption lattices of the distinct strata do not specify any particular 'lexical items' in the sense of lexicon entries. The types of the lattices and their interdependencies directly specify which combinations of features are *possible* in the language described — i.e., they represent the *linguistic potential* of the language. It is then a general property of languages that particular selections of linguistic features from the abstract linguistic potential can be grouped together to form (synchronically) arbitrary bundles: the features in a bundle will necessarily co-occur without being functionally predictable from each other. This information must be given in addition to the general statement of possible feature occurrences. It is this that is most usefully described by the term 'lexicon'. Importantly, entries here *only make sense* given the background that the general potential defines. The grammar and semantics thus offers a highly detailed hypothesis for the general organization of a 'lexicon' for each language described.

Several problematic features of lexicons are naturally resolved in such a view. For example, entries are clearly not restricted to the 'word', which is just one of the possible grammatical units: *any* feature combination can become fossilized in this way. This automatically covers phrasal lexicons, idiomatic expressions, frozen textual and rhetorical forms, dominant semantic patterns, lexical gender, lexical number, lexical case, etc. It also provides a natural link upwards to semantic organizations, since this is precisely the function of the clause grammar. The kind of semantics we then see is a *grammatical semantics*, which can very usefully complement more traditional notions of 'lexical semantics'. Moreover, of particular relevance to multilingual resource construction, it is the *clause* grammar where we see most commonalities across languages. Adopting organizations from the clause grammar and its semantic correlate thus also offers a scheme more likely to function multilingually and to support resource construction. For these reasons, I believe that it is now time for the two linguistic traditions of lexical semantics and grammatical semantics to share more of their results and insights.