

Abstract for "Structuring the Argument: A Multidisciplinary Workshop on the Mental Representation of Verbal Argument Structure" (Bachrach et al, Paris, 5th-7th Sept). 500 words.

The role of verb subcategorization frames and selectional preferences in sentence processing: an investigation using corpus-derived measures

Barry Devereux, Anna Korhonen, Paula Buttery, LK Tyler

A fundamental consequence of the temporal nature of speech is that temporary ambiguities arise where multiple candidate syntactic representations are consistent with the currently available input. Lexicalist theories propose that lexico-syntactic knowledge associated with constituent words results in activation of candidate parses during sentence comprehension (Marslen-Wilson et al. 1988; Tyler & Marslen-Wilson 1977; MacDonald et al. 1994). However, although it has been hypothesized that such knowledge reflects statistical data on words' usage in language, studies typically estimate lexico-syntactic information using either behavioural pre-tests or experimenter intuition (Shapiro et al. 1993; Connine et al. 1984; Novais-Santos et al. 2007). Here, we used corpus-derived statistical data to model two specific kinds of lexico-syntactic information associated with verbs, in order to investigate whether such information influences sentence processing.

Stimuli were spoken sentences containing potential local ambiguities where a verb participle functioned as either a gerund (e.g. "Understandably, *insulting* neighbours is not encouraged") or adjective ("Understandably, *insulting* neighbours are not respected"). We used the VALEX lexicon, derived from automatically-parsed cross-domain corpora (Korhonen et al. 2006), to obtain frequency distributions for each verb across subcategorization frames. We hypothesized that for verbs with a high probability of occurrence with noun-phrase direct object complements (e.g. "mark") there would be a preference for gerundive readings, because in such cases the following noun is likely to function as the verb's theme (e.g. "marking essays"), whereas for verbs with a low probability of occurrence with direct object complements there would be a stronger preference for adjectival readings (e.g. "yawning audiences").

Our second measure of lexico-syntactic knowledge reflected verb selectional preferences. Again using VALEX, we calculated probability distributions over WordNet for both the subject and direct object argument slots of each verb (following earlier approaches to modeling selectional preferences with WordNet, e.g. McCarthy 2001). We hypothesized that when the semantic class of the noun had a stronger association with the direct object slot than the subject slot (e.g. "cooking cabbages") there would be a preference for gerundive readings, whereas when the semantic class of the noun had a stronger association with the subject slot (e.g. "cooking husbands") there would be a stronger preference for adjectival readings. Importantly, the direct object preference measure and the

selectional preference measure were uncorrelated, indicating that they reflect different aspects of verbs' lexico-syntactic representations.

In Experiment 1, participants heard the sentences up to the end of the potentially ambiguous phrase and were instructed to write down plausible completions, which were then coded as being consistent with either gerundive or adjectival readings. In Experiment 2, participants heard the same incomplete sentences, followed by a disambiguating continuation word ("is" or "are") to which participants made acceptability judgments. We found that both the direct object preference measure and the selectional preference measure were predictive of both the continuation preferences in Experiment 1 and the acceptability judgment latencies in Experiment 2. These results support a lexicalist model of syntactic processing, where experience with verb lexico-syntactic behaviour (as revealed through corpus statistics) influences parsing preferences for local ambiguities during sentence processing.

Current length: 500 words