Sense-contingent lexical bias and its role for early parsing decisions

Recent research in sentence comprehension has provided both experimental evidence (cf, e.g., MacDonald 1994, Trueswell 1996) and evidence from connectionist modeling (cf., e.g., Kawamoto 1993) that lexically specific information plays a key role in syntactic ambiguity resolution. The present study offers a corpus-based approach to sentence processing that supports this view, but suggests that the probabilistic information most relevant for early parsing is situated at the level of individual word senses as opposed to the more general level of words.

One of the most frequently studied local syntactic ambiguities involves the alternation between nominal and sentential complements. In this ambiguity, a post-verbal NP cannot straightforwardly be interpreted with respect to the grammatical role it plays in the sentence. As illustrated in (1), it can function either as the direct object of the main verb, as in (1a), or, alternatively, the subject of the embedded clause as in (1b).

(1)

a. **Inspector Clousseau revealed [$_{NP}$ Dreyfuss's intentions].**

b. **Inspector Clousseau revealed [$_{S}$ [$_{NP}$ Dreyfuss's intentions] were indeed diabolic.]**

Early theories of sentence comprehension (e.g. Frazier and Fodor 1978, Frazier 1979) argued for a two-stage serial processing mechanism, which in its initial phase only uses syntactic information and general heuristics to guide early parsing decisions. In contrast, 'lexical guidance' accounts of sentence comprehension (cf., e.g., Ford et al. 1982, Mitchell 1987) assume that particular lexical items build up expectations for a particular parse. However, as example (2) illustrates, verbs usually have several senses, i.e. they can instantiate different semantic predicates which in turn may have different argument structure preferences.

(2)

a. Peter $_{VP}$ [$_{V}$ admitted$_1$ [$_{NP}$ **his ex-girlfriend**] [$_{PP}$ to the club]].

b. Peter $_{VP}$ [$_{V}$ admitted$_2$ [$_{S}$ [$_{NP}$ **his ex-girlfriend**] was hotter than his current one]].

c. Peter $_{VP}$ [$_{V}$ admitted$_2$ [$_{NP}$ **his error**]].

In its usage in (2a), the verb *admit* roughly means 'permit entry' and instantiates a predicate involving three arguments, which may be assigned the roles of AGENT, PATIENT, and GOAL. The crucial PATIENT argument is typically realized by an NP and, hence, a preference towards the syntactic form in (1a), i.e. the nominal complementation, is to be expected on semantic grounds. Furthermore, as (2b) and (2c) show, *admit* can also instantiate a cognitive or mental predicate, i.e. a two-place relation between a human agent and a proposition, and because propositions are typically expressed formally by means of the syntactic unit clause, the syntactic pattern in (2b) seems a natural continuation. Note, however, that the same verb sense can also occur with nominal complementations (see example 2c), arguably due to metonymic or metaphoric extension (cf., e.g., Lakoff 1987).

A recent experimental study (Hare et al. 2003) suggests that the relevant argument structure preferences, which crucially guide early parsing decisions, are situated at the level of these verb senses rather than at a more general verb level. The current paper, in accordance with this study, presents corpus evidence that verb senses drive the reconstruction of syntactic structure.

The analysis proceeds in two steps: First, I determine both the form-based and sense-contingent preferences of 20 verbs on the basis of a 17 million words sample of the BNC which is isomorphic to the British component of the ICE corpus (see Nelson 1996 for details). For the sense-contingent preferences, the semantic distinctions assumed in WordNet were used (cf. Fellbaum 1998) and all corpus data were coded manually (N_{token} ~ 5,000). To express the individual preferences, I apply a corpus-based method termed 'distinctive collexeme analysis' (Gries and Stefanowitsch 2004), which is a sub-type of a family of techniques geared to measure the interaction between lexical items and grammatical structures – or in Construction Grammar parlance, the interaction of two constructions of different levels of abstraction. The method usually employs association measures from statistical hypothesis testing to estimate the strength of attraction between the two (linguistic) units under investigation. Since some of the verb senses are quite infrequent, I make use of Fisher's exact test (cf. Agresti 1990) to estimate the preference. The results clearly indicate that the preferences associated with the individual verb sense strongly differ form the overall preferences of a given verb, hence predicting different structural expectations contingent on the verb sense instantiated.

Second, the results were compared with the data of a self-paced moving window

experiment reported in Hare et al. (2003). Specifically, a correlational analysis was performed between the results of the distinctive collexeme analysis, i.e. the computed associations scores, and the observed ambiguity effects from Hare et al's study, i.e. the individual reading time latencies at the critical position of the disambiguating region. The outcome of the correlational analysis is significant, which supports the experimental findings and thus provides additional evidence for the hypothesis that verb senses, rather than verb forms, play an important role in syntactic ambiguity resolution [tau = 0.0308; z = 0.2807; p = 0.048].

The findings of the study are consistent with Roland and Jurafsky's (2002) 'Lemma Argument Probability hypothesis', which states that word senses, or lemmas, bear the most relevant argument structure expectations, but are also compatible with accounts that conceive the locus of argument structure expectations to be on a more abstract level of semantic representation, say, semantically coherent verb classes.

In conclusion, the results of this present study suggest, on a theoretical level, that a general, verb specific conception of lexical bias is insufficient to characterize the effects associated with argument structure expectations during the process of sentence comprehension. Consequently, psychological models and experimental protocols using subcategorization preferences should take verb sense into account. From a methodological perspective, the study shows that the relevant preferences can be appropriately estimated by means of quantitative corpus-linguistic methodologies.

References:

Fellbaum, C. 1998**.** *WordNet: An electronic lexical database*. Cambridge: MIT Press.

Ford, M., Bresnan, J., and Kaplan, R.M. 1982. A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The Mental Representations of Grammatical Relations*. Cambridge, MA: MIT Press.

Frazier, L. 1979. On comprehending sentences: Syntactic parsing strategies (PhD thesis, University of Connecticut). West Bend, IN: Indiana University Linguistics Club.

Frazier, L., and Fodor, J.D. 1978. The sausage machine: A new two-stage parsing model. *Cognition*, 6,291-326

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48, 281-303

Kawamoto, A.H. 1993. Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, 32, 476-516

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories tell us about the nature of thought*. Chicago: University of Chicago Press.

MacDonald, M.C. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201

Mitchell, D.C. 1989. Verb guidance and lexical effects in ambiguity resolution. *Language and Cognitive Processes*, 4, 123-154

Nelson, Gerald (1996) The Design of the Corpus. In S. Greenbaum (ed.) *Comparing English Worldwide: The international corpus of English*, pp. 27-35. Oxford: Clarendon Press.

Roland, D., and Jurafsky, D. 2002. Verb sense and verb subcategorization probabilities. In Stevenson, Suzanne, and Paola Merlo (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*. Amsterdam: John Benjamins. 325-346.

Trueswell, J. 1996. The role of lexical frequency in syntactic ambiguity resolution. Journal of Memory and Language, 35, 566-585