Experience-Based HPSG

Emily M. Bender and Susanne Z. Riehemann
{*bender,sr*}*@csli.stanford.edu*

Berkeley Formal Grammar Conference
HPSG 2000

July 22nd, 2000

1 Introduction: Two views of competence

1.1 Mainstream view

- Chomsky claims psychological reality:

  > Statements about I-language ... are true or false ... The I-language L may be the one used by a speaker, but not the I-language L′, even if the two generate the same class of expressions... (Chomsky 1986:23)

- However:

  > In the transformational literature, it is customary to claim psychological reality for competence models on the one hand, but to deny, on the other, that they could be taken as process models. (Seuren 1982:4)

- Chomsky's insistence that the competence grammar is separate from performance is inconsistent with drawing on language acquisition (UG/LAD) or 'cognitive interfaces' (PF and LF) as sources of explanatory force.

1.2 HPSG

- Performance-plausible competence grammar (see also Bresnan 1978, Sag and Wasow 1999)

  > If the grammars offered by a linguistic theory are to be embedded into a theory of human language processing, then there is a variety of properties of language processing that might be expected to inform the design of grammar. (Pollard and Sag 1994:11)

- HPSG model of competence and performance:

  Competence *Knowledge of language*

  Performance *Use of knowledge of language*

- If a competence grammar is a description of an E-language, it's perfectly fine to limit its domain to the problem of determining all and only the sentences of the language.

- However, there is no evidence that, in the minds of actual speakers, there is a system of knowledge with only that function.

1.3 Usage-based models

- The usage-based approach, pioneered by Langacker (1987; 1990), provides an alternative view of competence that is fully compatible with HPSG.[1]

  > A usage-based model is one in which the speaker's linguistic system is fundamentally grounded in 'usage events': instances of a speaker's producing and understanding language. . . . [S]uch instances are the basis on which a speaker's linguistic system is formed. . . The linguistic system is built up from such . . . instances, only gradually abstracting more general representations. . . (Kemmer and Barlow 2000:ix)

---

[1]Bod (1998) and Neumann and Flickinger (1999) present computational systems that take advantage of storing pieces of previously encountered sentences. This bears a resemblance to certain aspects of the usage-based model and was part of the inspiration for the current paper.

- Explanatory force comes from bottom-up theory construction.

- Entrenchment/frequency can play a role in processing.

- Maximalism:

  > The grammar lists the full set of particular statements representing a speaker's grasp of linguistic convention, including those subsumed by general statements. ... Speakers do not necessarily forget the forms they already know once the rule is extracted, nor does the rule preclude their learning additional forms as established units. (Langacker 1987:46)

1.4   Outline of the rest of the talk

- Two new kinds of evidence for maximalist grammar.

- Discussion of how design features of HPSG actually fit right in.

2   Non-categorical constraints on variation

- Sociolinguistic variables are subject non-categorical grammatical constraints on the distribution of the variant forms.

- Speakers have knowledge of these constraints (Bender 2000), (Bender in press).

2.1   The variable: copula absence in AAVE

- African American Vernacular English (AAVE) allows copulaless sentences like (1b) alongside sentences like (1a).

  (1) a. She's teachin me piano.

   b. She teachin me piano.

- Copula absence/presence is affected by social and grammatical non-categorical constraints. For example, copula presence is more common with predicate nominals (2) than with *ing*-form verbs (1).

(2) a. She's my piano teacher.

    b. She my piano teacher.

2.2   The experiment

- Recordings of the sentences in (1) and (2) were presented to AAVE speakers and a control group.

- Listeners were asked to judge the speakers of the sentences on a number of personality characteristics (e.g., "How well educated does this person sound?")

- Ratings were compared to test for an effect of copula absence vs. presence, and of the predicate type.

2.3   The results

- To the AAVE speakers, copula presence sounded more educated than copula absence (among other effects).

- Copula presence before V+*ing* sounded even more educated than copula presence before a predicate nominal.

- This interaction between the social value of copula presence and the grammatical environment shows that speakers have knowledge of this non-categorical constraint.

- In a usage-based maximalist model, this could be represented with more specified types: i.e., lexical types for the copula with information about the complement and the social value.

3   Canonical forms of idioms

- Idioms and collocations occur in syntactically variable forms.

- However, they also have predominant canonical forms (Riehemann forthcoming).

3.1   Corpus data

- Corpus data shows a wide range of variation in the form of idioms.

    - Modification

        (3) Diana to *Spill Royal Beans* in Upcoming Interview

    - Passive

        (4) Even D'Amato has only theories on *what political strings may have been pulled.*

    - Split across clauses

        (5) Robert McNamara's new book justified *all the strings* Clinton *pulled* to keep out of Vietnam.

- However, canonical forms of idioms (in this case *spill the beans* and *pull strings*) account for roughly 70% of all corpus occurrences.

- Likewise, preliminary results indicate that collocations such as *bear the brunt of* NP appear in their canonical forms about 90% of the time.

3.2   Implications

- In order to handle the observed variability in the forms of idioms, it is necessary to have an underspecified representation for these idioms that is compatible with all of the forms.

- A minimalist grammar would only include this representation.

- However, the high level of canonicity for idioms and collocations represents knowledge that demands to be captured.

- The high level of canonicity doesn't follow from anything else:

    - Possible semantic explanation: the most frequent form for a non-idiom with a similar meaning to *spill the beans* is *reveal secrets*, not *reveal the secret*. In fact, *reveal secrets* accounts for only 10% of the forms, leaving no predominant form.

– Possible (un)markedness explanation: idioms vary enough in their forms (compare *spill the beans*, *pull strings*, *get a word in edgewise*) that it can't be the case that the most common form is simply unmarked grammatically.

- In a maximalist grammar, it is to be expected that speakers would have representations of canonical forms as subtypes of the more underspecified idiom types.

- Further evidence for including canonical form types comes from psycholinguistic studies (e.g., McGlone et al. (1994)) which have shown that idioms are processed faster when in canonical form than when in variant forms.

- Evidence that this subtype is part of the grammar:

  – Canonical form types make reference to (i.e., inherit from) many basic grammatical types (e.g., *head-complement-phrase*).
  – Where else could such knowledge reside?

4   Conclusion: Design features of HPSG

- Many of the design features of HPSG already seem to reflect a usage-based view of linguistic competence:

  – A crucial feature of the usage-based model is partial generalizations, and a grammar includes many levels of intermediate generalizations. This is already the hallmark of the HPSG type hierarchy.
  – The sign-based nature of HPSG and the inclusion of CONTEXT alongside CONTENT in the meaning of signs meshes well with the insistence in the usage-based literature on "the crucial role of context in the operation of the linguistic system" (Kemmer and Barlow 2000:xxi).
  – The declarative constraint-based nature of HPSG leads to a metaphor for determining grammaticality in which any sentence that can be built up by unifying objects in the grammar is in, all others are out. There are no well-formedness checks to be

performed at the end of a parse. This metaphor extends easily to the construction of sentences out of pieces that are larger or more specified than the maximally abstract types of a minimalist grammar.

- However, a usage-based view differs somewhat from current practice in HPSG. In particular, the grammar must be broadened to include certain overspecified types.

- These overspecified types form the connection between the grammar and usage. They are minimally abstracted from actual utterances.

References

Bender, Emily M. 2000. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence.* PhD thesis, Stanford University.

Bender, Emily. in press. Non-categorical constraints in perception. *Penn Working Papers in Linguistics* 7(1).

Bod, Rens. 1998. *Beyond Grammar: An Experience-Based Theory of Language.* Stanford, CA: CSLI.

Bresnan, Joan. 1978. A realistic transformational grammar. In M. Halle, J. Bresnan, and G. Miller (Eds.), *Linguistic Theory and Psychological Reality*, 1–59. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use.* New York: Praeger.

Kemmer, Suzanne, and Michael Barlow. 2000. Introduction: A usage-based conception of language. In M. Barlow and S. Kemmer (Eds.), *Usage-Based Models of Language*, vii–xxviii. Stanford: CSLI.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Vol I: Theoretical Prerequisites.* Stanford, CA: Stanford University Press.

Langacker, Ronald W. 1990. *Concept, Image, and Symbol: The Cognitive Basis of Grammar.* New York: Mouton de Gruyter.

McGlone, Matthew S., Sam Glucksberg, and Cristina Cacciari. 1994. Semantic productivity and idiom comprehension. *Discourse Processes* 17:167–190.

Neumann, Günter, and Dan Flickinger. 1999. Learning stochastic lexicalized tree grammars from HPSG. Technical report, DFKI.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago: Chicago University Press.

Riehemann, Susanne. forthcoming. *A Constructional Approach to Idioms and Word Formation*. PhD thesis, Stanford University.

Sag, Ivan A., and Thomas Wasow. 1999. *Synactic Theory: A Formal Introduction*. Stanford, CA: CSLI.

Seuren, Pieter A. M. 1982. Internal variability in competence. *Linguistische Berichte* 77:1–31.