HPSG – A Synopsis

Formal model of syntax developed by C. Pollard and I. Sag in the late 1980s. In essence a variant of Generalized Phrase Structure Grammar in which the structure of syntactic units is constrained directly or indirectly by the properties of lexical units as heads (1). E.g. the head of *Bill hates silence* is (in this treatment) *hates*. The lexical properties of *hates* are such that it combines with a following noun phrase: [*hates silence*]. By a process of unification this is in turn a verb phrase, and, since *hates* is a verb in the third person singular, the phrase as a whole is one which, as a non-lexical head, combines, again by a process of unification, with a subject in the third person singular.

As in other models which extend the role of the lexicon, there are either no or few specific rules. As in Generalized Phrase Structure Grammar, there is no operation, as in earlier or contemporary versions of transformational grammar, by which units are moved.

Head-driven phrase structure grammar (HPSG) is a highly lexicalized, non-derivational generative grammar theory developed by Carl Pollard and Ivan Sag.[1][2] It is the immediate successor to generalized phrase structure grammar. HPSG draws from other fields such as computer science (data type theory and knowledge representation) and uses Ferdinand de Saussure's notion of the sign. It uses a uniform formalism and is organized in a modular way which makes it attractive for natural language processing.

An HPSG grammar includes principles and grammar rules and lexicon entries which are normally not considered to belong to a grammar. The formalism is based on lexicalism. This means that the lexicon is more than just a list of entries; it is in itself richly structured. Individual entries are marked with types. Types form a hierarchy. Early versions of the grammar were very lexicalized with few grammatical rules (schema). More recent research has tended to add more and richer rules, becoming more like Construction Grammar .[3]

The basic type HPSG deals with is the sign. Words and phrases are two different subtypes of sign. A word has two features: *[PHON]* (the sound, the phonetic form) and *[SYNSEM]* (the syntactic and semantic information), both of which are split into subfeatures. Signs and rules are formalized as typed feature structures.

A Sample Grammar

HPSG generates strings by combining signs, which are defined by their location within a type hierarchy and by their internal feature structure, represented by attribute value matrices (AVMs). [4][5] Features take types or lists of types as their values, and these values may in turn have their own feature structure. Grammatical rules are largely expressed through the constraints signs place on one another. A sign's feature structure describes its phonological, syntactic, and semantic properties. In common notation, AVMs are written with features in upper case and types in italicized lower case. Numbered indices in an AVM represent token identical values.

In the simplified AVM for the word "walks" below, the verb's categorical information is divided into features that describe it (HEAD) and features that describe its arguments (VALENCE).

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \left\langle \text{`walks'} \right\rangle \\ \text{SYNSEM} & \begin{bmatrix} \textit{synsem} & \\ \text{CAT} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{verb} \\ \text{VALENCE} & \begin{bmatrix} \text{SUBJ} & \left\langle \begin{bmatrix} \textit{synsem} & \\ \text{CAT}\,|\,\text{HEAD} & \textit{noun} \\ \text{CONT} & \boxed{1} \begin{bmatrix} \textit{ref-index} & \\ \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sing} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMP} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{content} & \\ \text{WALKER} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

"Walks" is a sign of type *word* with a head of type *verb*. As an intransitive verb, "walks" has no complement but requires a subject that is a third person singular noun. The semantic value of the subject (CONTENT) is co-indexed with the verb's only argument (the individual doing the walking). The following AVM for "she" represents a sign with a SYNSEM value that could fulfill those requirements.

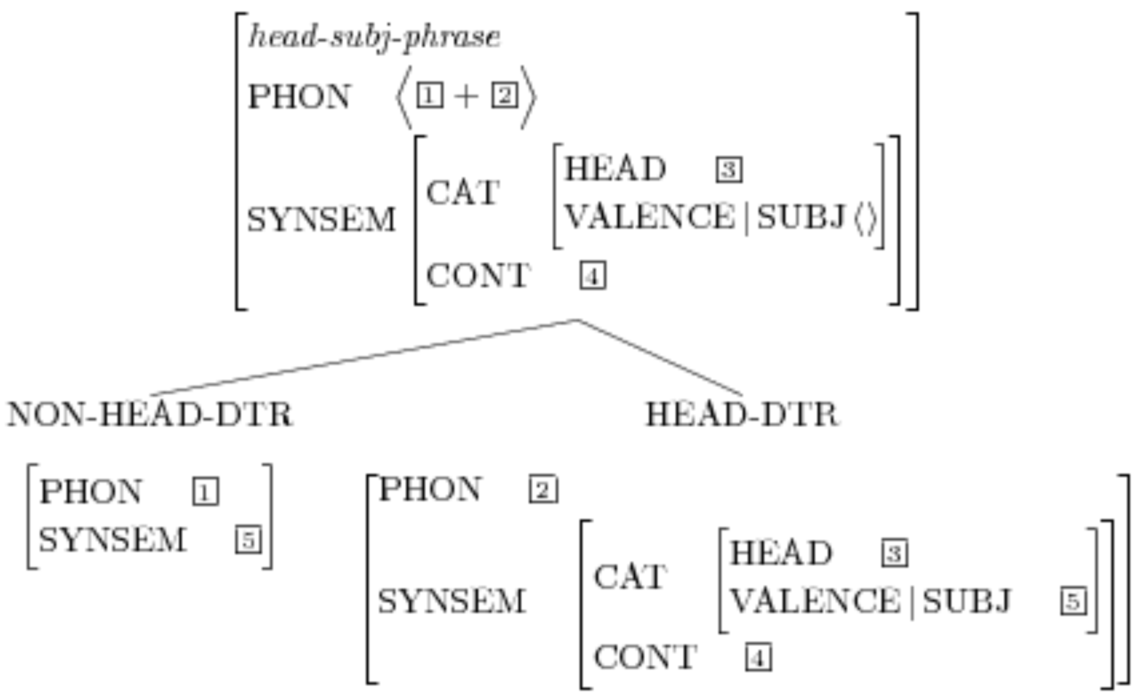
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \left\langle \text{`she'} \right\rangle \\ \text{SYNSEM} & \begin{bmatrix} \textit{synsem} & \\ \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{ref-index} & \\ \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sing} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Signs of type *phrase* unify with one or more children and propagate information upward. The following AVM encodes the immediate dominance rule for a *head-subj-phrase*, which requires two children: the head child (a verb) and a non-head child that fulfills the verb's SUBJ constraints.

$$
\begin{bmatrix}
\textit{head-subj-phrase} & \\
\text{PHON} & \left\langle \boxed{1} + \boxed{2} \right\rangle \\
\text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VALENCE} \mid \text{SUBJ} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \boxed{4} \end{bmatrix} \\
\text{HEAD-DTR} & \begin{bmatrix} \text{PHON} & \boxed{2} \\ \text{SS} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VALENCE} \mid \text{SUBJ} & \boxed{5} \end{bmatrix} \\ \text{CONT} & \boxed{4} \end{bmatrix} \end{bmatrix} \\
\text{NON-HEAD-DTR} & \begin{bmatrix} \text{PHON} & \boxed{1} \\ \text{SS} & \boxed{5} \end{bmatrix}
\end{bmatrix}
$$

The end result is a sign with a verb head, empty subcategorization features, and a phonological value that orders the two children.

Although the actual grammar of HPSG is composed entirely of feature structures, linguists often use trees to represent the unification of signs where the equivalent AVM would be unwieldy.



Implementations

Various parsers based on the HPSG formalism have been written and optimizations are currently being investigated. An example of a system analyzing German sentences is provided by the Freie Universität Berlin.[6] In addition the Grammar Group of the Freie Universität Berlin provides open source grammars that were implemented in the **TRALE** system. Currently there are grammars for German,[7] Mandarin Chinese,[8] Maltese,[9] and Persian[10] that share a common core and are publicly available. For Dutch, the wide-coverage dependency parser **Alpino** [11] has been developed at the University of Groningen.

Large HPSG grammars of various languages are being developed in the Deep Linguistic Processing with HPSG Initiative (DELPH-IN).[12] Wide-coverage grammars of German,[13] English [14] and Japanese [15] are available under an open-source license. These grammars can be used with the open-source HPSG systems **LKB** and **PET**. DELPH-IN grammars can typically be used for both parsing and generation. Treebanks also distributed by DELPH-IN are being used to develop and test the grammars, as well as train ranking models to decide on plausible interpretations when parsing (or realizations when generating).

Enju is a freely available wide-coverage probabilistic HPSG parser for English developed by the Tsujii Laboratory at The University of Tokyo in Japan. Its robustness sets it apart from most other HPSG parsers.[16]

References

1. ^ Pollard, Carl, and Ivan A. Sag. 1987. Information-based syntax and semantics. Volume 1. Fundamentals. CLSI Lecture Notes 13.
2. ^ Pollard, Carl; Ivan A. Sag. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
3. ^ Sag, Ivan A. 1997. English Relative Clause Constructions. Journal of Linguistics . 33.2: 431-484
4. ^ Pollard, Carl; Ivan A. Sag. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
5. ^ Sag, Ivan A.; Thomas Wasow; & Emily Bender. (2003). *Syntactic theory: a formal introduction*. 2nd ed. Chicago: University of Chicago Press.
6. ^ The Babel-System: HPSG Interactive
7. ^ Berligram
8. ^ Chinese
9. ^ Maltese
10. ^ Persian
11. ^ Alpino
12. ^ DELPH-IN: Open-Source Deep Processing
13. ^ Berthold Crysmann
14. ^ English Resource Grammar and Lexicon
15. ^ JacyTop - Deep Linguistic Processing with HPSG (DELPH-IN)
16. ^ Tsuji Lab: Enju parser home page (retrieved Nov 24, 2009)

Further reading

- Carl Pollard, Ivan A. Sag (1987): *Information-based Syntax and Semantics. Volume 1: Fundamentals*. Stanford: CSLI Publications.
- Carl Pollard, Ivan A. Sag (1994): *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press. ([1])
- Ivan A. Sag, Thomas Wasow, Emily Bender (2003): *Syntactic Theory: a formal introduction, Second Edition*. Chicago: University of Chicago Press. ([2])
- Levine, Robert D.; W. Detmar Meurers (2006). "Head-Driven Phrase Structure Grammar: Linguistic Approach, Formal Foundations, and Computational Realization". In Keith Brown. *Encyclopedia of Language and Linguistics* (second edition ed.). Oxford: Elsevier.

External links

- Stanford HPSG homepage – includes on-line proceedings of an annual HPSG conference
- Ohio State HPSG homepage
- DELPH-IN network for HPSG grammar development
- Basic Overview of HPSG
- Comparison of HPSG with alternatives, and a historical perspective
- Bibliography of HPSG Publications
- LaTeX package for drawing AVMs – includes documentation