Closer to the Truth: A New Model Theory for HPSG

Frank Richter

Eberhard Karls Universität Tübingen

August 17th, 2007

The Plot

1. HPSG for the description of languages
2. Shortcomings: Imprecise grammars
3. Normal form HPSG grammars
4. Three model theories and philosophical views
    4.1 King 1999
    4.2 Pollard & Sag 1994
    4.3 Pollard 1999
5. A new proposal
6. Conclusion

The Signature Σ_1

top
- *sign* PHON *list*, CAT *cat*
  - *phrase* H-DTR *sign*, NH-DTR *sign*
  - *word*
- *list*
  - *nelist* FIRST *top*, REST *list*
  - *elist*
- *cat* HEAD *head*, SUBCAT *list*
- *head*
  - *verb*
  - *noun*
- *phonstring*
  - *uther*
  - *walks*

append/3

The Theory θ_1

- WORD PRINCIPLE:

$$[word] \rightarrow \left(\begin{bmatrix} \text{PHON} & \langle \textit{uther} \rangle \\ \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{SUBCAT} & \textit{elist} \end{bmatrix} \end{bmatrix} \vee \begin{bmatrix} \text{PHON} & \langle \textit{walks} \rangle \\ \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SUBCAT} & \left\langle \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{SUBCAT} & \textit{elist} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right)$$

- ID PRINCIPLE:

$$[phrase] \rightarrow \begin{bmatrix} \text{CAT SUBCAT} & \textit{elist} \\ \text{H-DTR CAT SUBCAT} & \langle \boxed{1} \rangle \\ \text{NH-DTR CAT} & \boxed{1} \end{bmatrix}$$

- HEAD FEATURE PRINCIPLE:

$$[phrase] \rightarrow \begin{bmatrix} \text{CAT HEAD} & \boxed{1} \\ \text{H-DTR CAT HEAD} & \boxed{1} \end{bmatrix}$$

- CONSTITUENT ORDER PRINCIPLE:

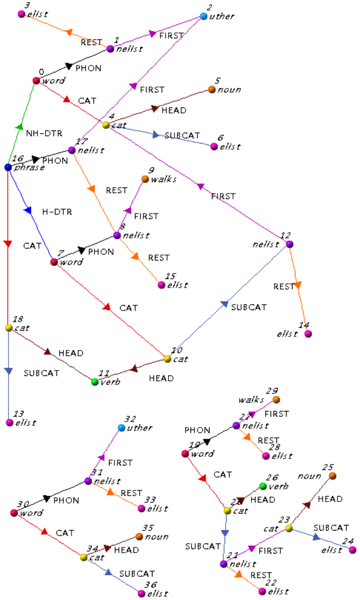
$$[phrase] \rightarrow \left(\begin{bmatrix} \text{PHON} & \boxed{3} \\ \text{H-DTR PHON} & \boxed{2} \\ \text{NH-DTR PHON} & \boxed{1} \end{bmatrix} \wedge \text{append}(\boxed{1}, \boxed{2}, \boxed{3}) \right)$$

The Theory θ_1 (completed)

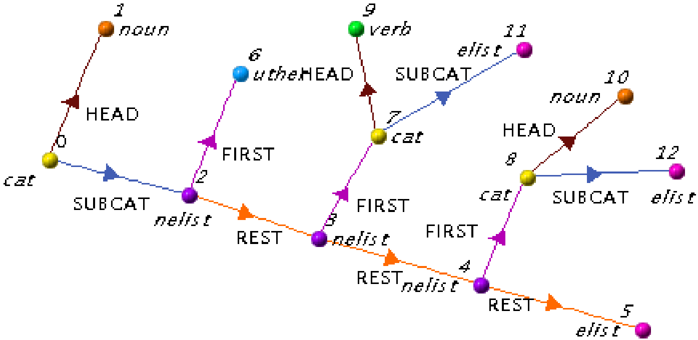
- Append Principle:

$$\forall \boxed{1} \forall \boxed{2} \forall \boxed{3} \left(\begin{array}{l} \text{append}(\boxed{1}, \boxed{2}, \boxed{3}) \leftrightarrow \\ \left(\begin{array}{l} \left(\boxed{1}[\textit{elist}] \wedge \boxed{2}[\textit{list}] \wedge \boxed{2} = \boxed{3} \right) \vee \\ \exists \boxed{4} \exists \boxed{5} \exists \boxed{6} \left(\begin{array}{l} \boxed{1} \langle \boxed{4} \mid \boxed{5} \rangle \wedge \boxed{3} \langle \boxed{4} \mid \boxed{6} \rangle \\ \wedge \ \text{append}(\boxed{5}, \boxed{2}, \boxed{6}) \end{array} \right) \end{array} \right) \end{array} \right)$$

The Intended Result



A Stranded Monster



$$\text{append} = \left\{ \begin{array}{l} \langle 2,5,2\rangle, \langle 3,5,3\rangle, \langle 4,5,4\rangle, \langle 5,2,2\rangle, \langle 5,3,3\rangle, \\ \langle 5,4,4\rangle, \langle 5,5,5\rangle, \langle 5,11,11\rangle, \langle 5,12,12\rangle, \\ \langle 11,2,2\rangle, \langle 11,3,3\rangle, \langle 11,4,4\rangle, \langle 11,5,5\rangle, \\ \langle 11,11,11\rangle, \langle 11,12,12\rangle, \langle 12,2,2\rangle, \langle 12,3,3\rangle, \\ \langle 12,4,4\rangle, \langle 12,5,5\rangle, \langle 12,11,11\rangle, \langle 12,12,12\rangle \end{array} \right\}$$

The Plot

1. HPSG for the description of languages
2. Shortcomings: Imprecise grammars
3. Normal form HPSG grammars
4. Three model theories and philosophical views
   4.1 King 1999
   4.2 Pollard & Sag 1994
   4.3 Pollard 1999
5. A new proposal
6. Conclusion

Normal Form Grammars 1

We add to each HPSG grammar

- Σ a sort hierarchy of signs which distinguishes unembedded signs from embedded signs,
- Σ an attribute, appropriate to each sort, which articulates the insight that each entity in the linguistic universe has the property of belonging to an unembedded sign,
- θ a principle which requires that each entity be a component of an unembedded sign,
- θ a principle which requires the uniqueness of unembedded sign entities in connected configurations of entities, and, finally,
- θ a principle which formulates the weak extensionality of *elist* entities.

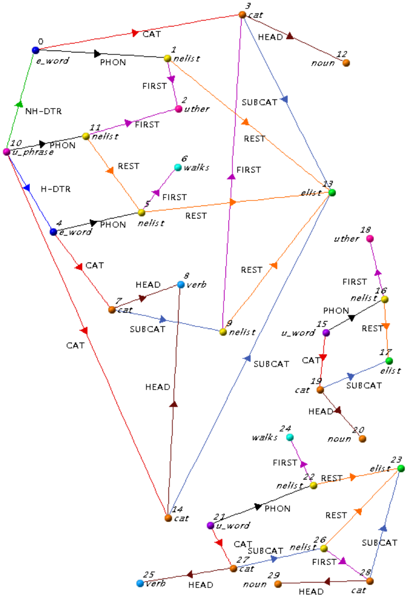
Normal Form Grammars 2

Normal form extension Σ_2 of signature Σ_1:

```
top  EMBEDDED  u_sign
    sign  PHON  list
          CAT   cat
        e_sign
            e_word
            e_phrase
        u_sign
            u_word
            u_phrase
        word
            e_word
            u_word
        phrase  H-DTR   e_sign
                NH-DTR  e_sign
            e_phrase
            u_phrase
    list
```

...
`component/2`

The Intended Model Fixed



$$\texttt{append} = \left\{ \begin{array}{l} \langle 1, 13, 1\rangle, \langle 1, 5, 11\rangle, \langle 13, 1, 1\rangle, \langle 13, 13, 13\rangle, \\ \langle 13, 5, 5\rangle, \langle 13, 9, 9\rangle, \langle 13, 11, 11\rangle, \\ \langle 5, 13, 5\rangle, \langle 9, 13, 9\rangle, \langle 11, 13, 11\rangle, \\ \langle 16, 17, 16\rangle, \langle 17, 16, 16\rangle, \langle 17, 17, 17\rangle, \\ \langle 22, 23, 22\rangle, \langle 23, 22, 22\rangle, \langle 23, 23, 23\rangle, \\ \langle 23, 26, 26\rangle, \langle 26, 23, 26\rangle \end{array} \right\}$$

$$\texttt{component} = \begin{array}{l} \{0, 1, \ldots, 14\} \times \{0, 1, \ldots, 14\} \ \cup \\ \{15, 16, \ldots, 20\} \times \{15, 16, \ldots, 20\} \ \cup \\ \{21, 22, \ldots, 29\} \times \{21, 22, \ldots, 29\} \end{array}$$

The Plot

1. HPSG for the description of languages
2. Shortcomings: Imprecise grammars
3. Normal form HPSG grammars
4. Three model theories and philosophical views
   4.1 King 1999
   4.2 Pollard & Sag 1994
   4.3 Pollard 1999
5. A new proposal
6. Conclusion

Direct Denotation (King 1999)

Central concern:

What does it mean for a grammar to be true of a natural language?

Illustration of AFS (Pollard & Sag 1994)

$\mathbb{A}_{\text{Uther}} = \langle \beta_{\text{U}}, \varrho_{\text{U}}, \lambda_{\text{U}}, \xi_{\text{U}} \rangle$ with

$$\beta_{\text{U}} = \left\{ \begin{array}{l} \epsilon, \text{PHON}, \text{PHON REST}, \text{PHON FIRST}, \\ \text{CAT}, \text{CAT SUBCAT}, \text{CAT HEAD} \end{array} \right\},$$

$$\varrho_{\text{U}} = \left\{ \begin{array}{l} \langle \epsilon, \epsilon \rangle, \langle \text{PHON}, \text{PHON} \rangle, \langle \text{CAT}, \text{CAT} \rangle, \\ \langle \text{PHON FIRST}, \text{PHON FIRST} \rangle, \\ \langle \text{PHON REST}, \text{PHON REST} \rangle, \\ \langle \text{PHON REST}, \text{CAT SUBCAT} \rangle, \\ \langle \text{CAT SUBCAT}, \text{PHON REST} \rangle, \\ \langle \text{CAT SUBCAT}, \text{CAT SUBCAT} \rangle, \\ \langle \text{CAT HEAD}, \text{CAT HEAD} \rangle \end{array} \right\},$$

$$\lambda_{\text{U}} = \left\{ \begin{array}{l} \langle \epsilon, \textit{u\_word} \rangle, \langle \text{PHON}, \textit{nelist} \rangle, \\ \langle \text{PHON REST}, \textit{elist} \rangle, \\ \langle \text{CAT SUBCAT}, \textit{elist} \rangle, \\ \langle \text{PHON FIRST}, \textit{uther} \rangle, \langle \text{CAT}, \textit{cat} \rangle, \\ \langle \text{CAT HEAD}, \textit{noun} \rangle \end{array} \right\},$$

$$\xi_{\text{U}} = \left\{ \begin{array}{l} \langle \texttt{append}, \text{PHON}, \text{PHON REST}, \text{PHON} \rangle, \\ \langle \texttt{append}, \text{PHON REST}, \text{PHON}, \text{PHON} \rangle, \\ \langle \texttt{append}, \text{PHON}, \text{CAT SUBCAT}, \text{PHON} \rangle, \\ \langle \texttt{append}, \text{CAT SUBCAT}, \text{PHON}, \text{PHON} \rangle \end{array} \right\} \cup \left\{ \langle \texttt{append}, \pi_1, \pi_2, \pi_3 \rangle \;\middle|\; \begin{array}{l} \pi_1, \pi_2, \pi_3 \in \\ \left\{ \begin{array}{l} \text{PHON REST}, \\ \text{CAT SUBCAT} \end{array} \right\} \end{array} \right\}$$

$$\cup \left\{ \langle \texttt{component}, \pi_1, \pi_2 \rangle \;\middle|\; \begin{array}{l} \pi_1 \in \beta_{\text{U}}, \pi_2 \in \beta_{\text{U}}, \& \\ \pi_1 = \pi_2 \text{ or} \\ \pi_2 \text{is a prefix of } \pi_1 \end{array} \right\}$$

Illustration of AFS (Pollard & Sag 1994)

The PHON FIRST reduct of $\mathbb{A}_{\text{Uther}}$:

$\beta_{\text{PF}} = \{\epsilon\}$,
$\varrho_{\text{PF}} = \{\langle\epsilon, \epsilon\rangle\}$,
$\lambda_{\text{PF}} = \{\langle\epsilon, \textit{uther}\rangle\}$, and
$\xi_{\text{PF}} = \{\langle \texttt{component}, \epsilon, \epsilon\rangle\}$.

All reducts represented as graphs:

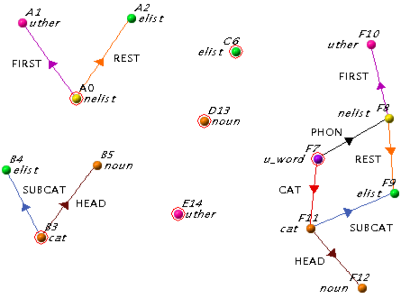


Illustration of MI (Pollard 1999)

Extraction of canonical representatives from exhaustive models using an abstraction function:

- $\langle u_{15}, \langle U_{15}, S_{15}, A_{15}, R_{15} \rangle \rangle$ (*u_word*)
- $\langle u_{16}, \langle U_{16}, S_{16}, A_{16}, R_{16} \rangle \rangle$ (*nelist*)
- $\langle u_{17}, \langle U_{17}, S_{17}, A_{17}, R_{17} \rangle \rangle$ (*elist*)
- $\langle u_{18}, \langle U_{18}, S_{18}, A_{18}, R_{18} \rangle \rangle$ (*uther*)
- $\langle u_{19}, \langle U_{19}, S_{19}, A_{19}, R_{19} \rangle \rangle$ (*cat*)
- $\langle u_{20}, \langle U_{20}, S_{20}, A_{20}, R_{20} \rangle \rangle$ (*noun*)

The Plot

A New Proposal

- Connected configurations in interpretations
- Maximal connected configurations in interpretations
- Minimal exhaustive models

Concluding Observations

- Merits of the framework: Meaningful discussion of subtle philosophical distinctions
- Alternatives within one single formalism
- HPSG as framework with sophisticated theory of the modelling structures of grammars
- Relevance for MTS in general?

The End

Thank you!

Extras: Encoding the Normal Form Extension

Normal form extension θ_2 of theory θ_1:

- U-SIGN COMPONENT CONDITION:

  $\forall\boxed{1}\,(\boxed{1}\,[top] \rightarrow \exists\boxed{2}\,\mathtt{component}(\boxed{1},\ \boxed{2}\,[u\_sign]\,))$

- UNIQUE U-SIGN CONDITION:

  $\forall\boxed{1}\ \forall\boxed{2}\,((\ \boxed{1}\,[u\text{-}sign]\ \wedge\ \boxed{2}\,[u\text{-}sign]\ ) \rightarrow\ \boxed{1} = \boxed{2})$

- UNIQUE EMPTY LIST CONDITION:

  $\forall\boxed{1}\ \forall\boxed{2}\,((\ \boxed{1}\,[elist]\ \wedge\ \boxed{2}\,[elist]\ ) \rightarrow\ \boxed{1} = \boxed{2})$

- COMPONENT PRINCIPLE:

$$\forall\boxed{1}\forall\boxed{2}\ \left(\begin{array}{l}\mathtt{component}(\boxed{1},\boxed{2}) \leftrightarrow \\ \left(\begin{array}{l}\boxed{1} = \boxed{2}\ \vee \\ \bigvee_{\alpha\in\mathcal{A}}\exists\boxed{3}\,(\ \boxed{2}\,[\alpha\ \boxed{3}] \wedge \mathtt{component}(\boxed{1},\boxed{3}))\end{array}\right)\end{array}\right)$$