Linguistic Data Structrures: Theoretical, Methodological, Descriptive Issues

Three kinds of "underspecified semantics" in HPSG

Frank Richter and Manfred Sailer
Universität Tübingen
{fr,mf}@sfs.nphil.uni-tuebingen.de

December 19, 2000

1 Introduction

1.1 Syntax and Semantics in HPSG

Descriptions and their denotation:

- A grammar consists of a **signature** and a **theory**
- A **model** of a grammar is a collection of objects whose components are configured according to the signature; and all objects are described by every description in the theory.
- An **exhaustive model** contains "instances" of all objects that are licensed by the theory.

HPSG is *sign-based.* This means that:

- every sign contains both syntactic and semantic structure (Pollard and Sag, 1994):

$$\begin{bmatrix} \textit{sign} & \\ \textsc{phonology} & \textbf{phonological structure} \\ \textsc{synsem} & \begin{bmatrix} \textsc{local} & \begin{bmatrix} \textsc{category} & \textbf{syntactic structure} \\ \textsc{content} & \textbf{semantic structure} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

- thus, for a sentence which has only one syntactic analysis but which is semantically ambiguous, there are two distinct signs that model the sentence.

(1) a. Every student read some book.
b. $\forall x[\mathsf{student}'(x) \rightarrow \exists y[\mathsf{book}'(y) \wedge \mathsf{read}'(x,y)]]$
c. $\exists y[\mathsf{book}'(y) \wedge \forall x[\mathsf{student}'(x) \rightarrow \mathsf{read}'(x,y)]]$

General assumptions made in this talk:

- We take *Relational Speciate Re-entrant Language* (RSRL, Richter et al. (1999); Richter (2000)) as the description language for HPSG grammars.
- We assume that the CONTENT value of a sign is a **logical form** (Richter and Sailer, 1999a,b; Sailer, 2000). In particular, we assume a language like IL (Montague, 1974) or Ty2 (Gallin, 1975).

(2) A description that denotes the term $\lambda x_e.\mathsf{student}'_{et}(x_e)$:

$$\begin{bmatrix} \textit{abstraction} & \\ \text{\textsc{type}} & \begin{bmatrix} \textit{c-type} & \\ \text{\textsc{in}} & \boxed{1}\ \textit{entity} \\ \text{\textsc{out}} & \boxed{2}\ \textit{truth} \end{bmatrix} \\ \text{\textsc{var}} & \boxed{3} \begin{bmatrix} \textit{var} & \\ \text{\textsc{type}} & \boxed{1} \end{bmatrix} \\ \text{\textsc{arg}} & \begin{bmatrix} \textit{application} & \\ \text{\textsc{func}} & \begin{bmatrix} \textit{student} & \\ \text{\textsc{type}} & \begin{bmatrix} \textit{c-type} & \\ \text{\textsc{in}} & \boxed{1} \\ \text{\textsc{out}} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{\textsc{arg}} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

1.2 Underspecification

The most natural notion of "*underspecification*" in HPSG is *underspecification at the description level*

- *underspecification in the lexicon*: a lexical entry is a description. A lexical entry is underspecified if there is some attribute whose value is not fully determined by the lexical entry. For example, for auxiliary verbs in English, the grammar in Pollard and Sag (1994) leaves the value of the INVERTED attribute underspecified. In every sentence, however, the value of this attribute is either *plus* or *minus*.
- In the grammar of Pollard and Sag (1994) sentence (1) receives a single syntactic analysis. The SEMANTICS PRINCIPLE leaves some freedom as to the place where quantifiers are inserted into the CONTENT value of a sign. In this sense, the SEMANTICS PRINCIPLE is underspecified.

Three criteria for semantic systems for HPSG:

1. **underspecified denotation**: the CONTENT value represents an underspecified semantic term.
2. **indirect representation**: meta-variables occur in the CONTENT value.
3. **discontinuous representation**: the semantic contribution of words is not a single term but a collection of terms; these terms are usually discontinually distributed over the logical form of bigger syntactic units.

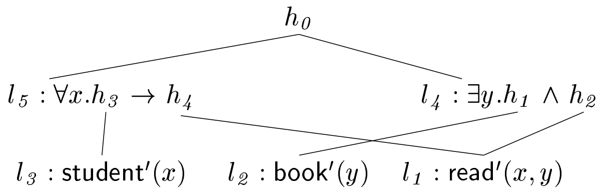
Table 1: Classification of the semantic systems discussed in this presentation

Approach	underspecified den.	indirect repr.	discontinuous repr.
Ty2U	+	+	+
Ty2U^P	-	+	+
LRS	-	-	+
LF-Ty2	-	-	-

2 Underspecified Denotation: Ty2U (Richter and Sailer, 1999c)

2.1 Bos (1995)

(3) An underspecified representation of the sentence *every student read some book.*:



Underspecified Representation:

(4) a. A set of holes:

$\{h_0, h_1, h_2, h_3, h_4\}$,

b. A set of labelled formulae:

$\{l_1 : \mathsf{read}'(x,y),\quad l_2 : \mathsf{book}'(y),\quad l_3 : \mathsf{student}'(x), l_4 : \exists y.h_1 \wedge h_2,\quad l_5 : \forall x.h_3 \to h_4\}$

c. A set of constraints:

$\{l_1 \leq h_0,\quad l_2 \leq h_0,\quad l_4 \leq h_0,\quad l_2 \leq h_1,\quad l_1 \leq h_2,\quad l_3 \leq h_0,\quad l_5 \leq h_0,\quad l_3 \leq h_3,\quad l_1 \leq h_4\}$

Plugging: bijection from holes to labels

(5) a. $\forall x.[\mathsf{student}'(x) \to \exists y.[\mathsf{book}'(y) \wedge \mathsf{read}'(x,y)]]$
$P1 = \{h_0 = l_5,\quad h_1 = l_2,\quad h_2 = l_1,\quad h_3 = l_3,\quad h_4 = l_4\}$

b. $\exists y.[\mathsf{book}'(y) \wedge \forall x.[\mathsf{student}'(x) \to \mathsf{read}'(x,y)]]$
$P2 = \{h_0 = l_3,\quad h_1 = l_2,\quad h_2 = l_5,\quad h_3 = l_4,\quad h_4 = l_1\}$

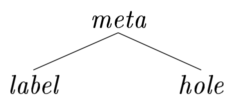
2.2 A Sketch

Signature for **underspecified representations**:

(6) a. *undersp.-repr*

undersp.-repr	HOLES	*list(hole)*
	LABELS	*list(label)*
	CONSTR(AINTS)	*list(constraint)*
	MAIN	*label*
	TOP	*hole*

b.



c.

constraint	A1	*meta*
	A2	*meta*

Relevant parts of some lexical entries:

(7) a. $$\begin{bmatrix} \textsc{phon}\ \langle \textit{read} \rangle & \\ \textsc{synsem loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head} & \textit{verb} \\ \textsc{subcat} & \langle \mathrm{NP},\ \mathrm{NP} \rangle \end{bmatrix} \\ \textsc{content} & \begin{bmatrix} \textit{ur} & \\ \textsc{holes} & \langle h_0 \rangle \\ \textsc{labels} & \langle \boxed{1} \lambda y \lambda x.\mathsf{read}'(x,y) \rangle \\ \textsc{constr} & \langle \boxed{1} \leq h_0 \rangle \\ \textsc{main} & \boxed{1} \\ \textsc{top} & h_0 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

b. $$\begin{bmatrix} \text{PHON} \ \langle \textit{book} \rangle \\ \text{SYNSEM LOC} \begin{bmatrix} \text{CAT HEAD} \ \textit{noun} \\ \text{CONTENT} \begin{bmatrix} \textit{ur} \\ \text{HOLES} & \langle h_0 \rangle \\ \text{LABELS} & \left\langle \boxed{2} \lambda y.\mathsf{book}'(y) \right\rangle \\ \text{CONSTR} & \left\langle \boxed{2} \leq h_0 \right\rangle \\ \text{MAIN} & \boxed{2} \\ \text{TOP} & h_0 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

c. $$\begin{bmatrix} \text{PHON} \ \langle \textit{student} \rangle \\ \text{SYNSEM LOC} \begin{bmatrix} \text{CAT HEAD} \ \textit{noun} \\ \text{CONTENT} \begin{bmatrix} \textit{ur} \\ \text{HOLES} & \langle h_0 \rangle \\ \text{LABELS} & \left\langle \boxed{3} \lambda x.\mathsf{student}'(x) \right\rangle \\ \text{CONSTR} & \left\langle \boxed{3} \leq h_0 \right\rangle \\ \text{MAIN} & \boxed{3} \\ \text{TOP} & h_0 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

d. $$\begin{bmatrix} \text{PHON} \ \langle \textit{some} \rangle \\ \text{SYNSEM LOC} \begin{bmatrix} \text{CAT HEAD} \ \textit{determiner} \\ \text{CONTENT} \begin{bmatrix} \textit{ur} \\ \text{HOLES} & \left\langle h_0, h_1, h_2 \right\rangle \\ \text{LABELS} & \left\langle \boxed{4} \exists y[h_1 \wedge h_2] \right\rangle \\ \text{CONTR} & \left\langle \boxed{4} \leq h_0 \right\rangle \\ \text{MAIN} & \boxed{4} \\ \text{TOP} & h_0 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

e. $$\begin{bmatrix} \text{PHON} \ \langle \textit{every} \rangle \\ \text{SYNSEM LOC} \begin{bmatrix} \text{CAT HEAD} \ \textit{determiner} \\ \text{CONTENT} \begin{bmatrix} \textit{ur} \\ \text{HOLES} & \left\langle h_0, h_3, h_4 \right\rangle \\ \text{LABELS} & \left\langle \boxed{5} \forall y[h_3 \rightarrow h_4] \right\rangle \\ \text{CONTR} & \left\langle \boxed{5} \leq h_0 \right\rangle \\ \text{MAIN} & \boxed{5} \\ \text{TOP} & h_0 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(8) The SEMANTICS PRINCIPLE:

In each phrase:

1. HOLES: is a concatenation of the values at the daughters
2. LABELS: is a concatenation of the values at the daughters
3. MAIN: is identical to the head daughter's value
4. TOP: is identical at the mother and the daughters
5. CONSTR: contains exactly
   (a) all elements of the daughters' CONSTR lists,
   (b) in *head-adjunct-phrase*:
   head's main $\leq$ adjunct's main
   adjunct's main $\leq$ top
   (c) in *head-complement-phrase*:
   head's main $\leq$ complement's nucleus
   complement's main $\leq$ top
   (d) in a nominal projection:
   head's main $\leq$ quantifier's restriction

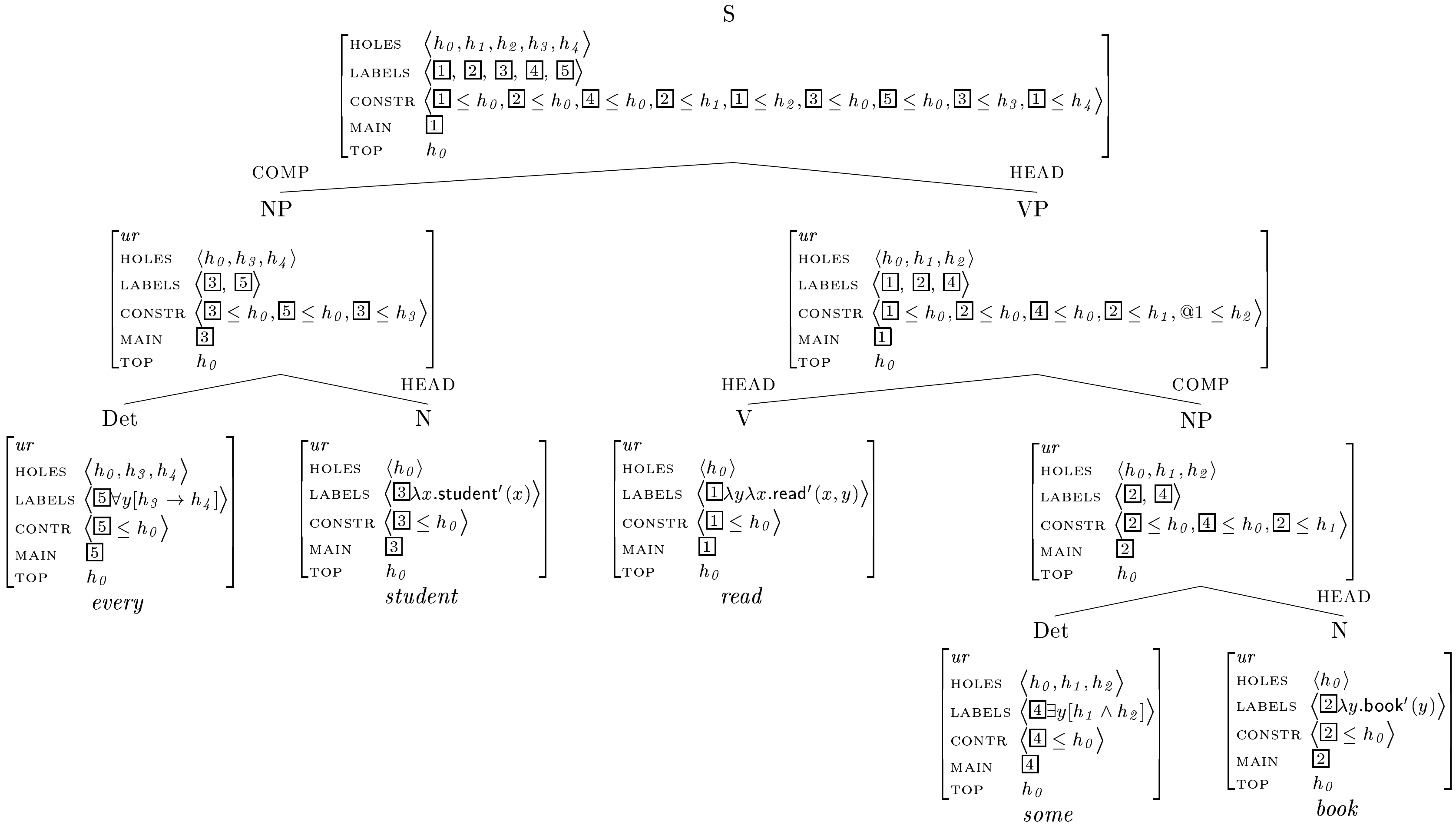


Figure 1: The analysis of the sentence *Every student read some book.*:

(9) a. description of the CONTENT value for the sentence *Every student read some book*:

$$\begin{bmatrix} \text{HOLES} & \langle h_0, h_1, h_2, h_3, h_4 \rangle \\ \text{LABELS} & \langle \boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5} \rangle \\ \text{CONSTR} & \langle \boxed{1} \leq h_0, \boxed{2} \leq h_0, \boxed{4} \leq h_0, \boxed{2} \leq h_1, \boxed{1} \leq h_2, \boxed{3} \leq h_0, \boxed{5} \leq h_0, \boxed{3} \leq h_3, \boxed{1} \leq h_4 \rangle \\ \text{MAIN} & \boxed{1} \\ \text{TOP} & h_0 \end{bmatrix}$$

b. This describes an underspecified representation which is like (4).

c. This underspecified representation has the pluggings in (5)

2.3 Problematic Data

In Ty2U scope resolution is not part of the grammar. Therefore, it is impossible to exclude a sentence whose ungrammaticality is due to conflicting scoping requirements.

Example 1: interrogatives in German (Beck, 1996)

(10) a. * Wann hat niemand wem geholfen?
when has nobody whom helped

b. Wann hat wem niemand geholfen?
when has whom nobody helped

'When did nobody help whom'
$\lambda p.\exists t \exists x[p = \lambda w.\neg\exists y[y \text{ helps } x \text{ at } t]]$

Assumptions:

1. Assume a quantificational analysis of interrogatives, i.e.,
wer (*who*) $\rightsquigarrow$ $\lambda P \lambda Q \lambda p.\exists x[P_w(x) \wedge p = \lambda w.Q_w(x)]$.
2. Assume that the relative scope of a negation and a quantifier is (largely) determined by word order.
3. The quantifier contributed by the in-situ interrogative pronoun *wem* ($\exists y$) must outscope the interrogative operator ($p = \lambda w.\phi$).

→ With (1) and (2) it follows that
(i) the negation must be in the scope of the fronted interrogative pronoun, and (ii) the quantifier contributed by the in-situ interrogative must be in the scope of the negation.

With (3) it follows that the quantifier contributed by *wem* must outscope the interrogative operator.

Thus, we get a contradiction for (10a) but nor for (10b).

Example 2: n-words in Polish (Przepiórkowski and Kupść, 1999)
In Polish, the preverbal particle *nie* is systematically ambiguous between eventuality negation and non-eventuality negation (pleonastic or other). N-words such as *nikt* (*no-one*) must co-occur with the particle *nie*, but can do so only in its eventuality negation reading.

(11) a. Omal jej nie przeróciłem.
almost her NM I overturned

'I almost knocked her over.'

b. ?* Omal nikogo nie przeróciłem.
almost nobody not I overturned

2.4 Technical Difficulties

- What kind of constraints are necessary? In Bos (1995) only constraints of the kind *"is subterm of/is in the scope of"* are possible. In Egg (1998) there are constraints of the forms *"is possibly in the immediate scope of"* and *"cannot possibly be in the immediate scope of"*. In Frank and Reyle (1995) there are even constraints of the form "$l_1 \leq l_2 \Rightarrow l_3 \leq l_4$".

- It is hard to express more complex constraints:

  (12) In a sentence S, if one in-situ interrogative pronoun takes scope over clause C then so does every clausemate in-situ interrogative pronoun.

  (13) a. Wer weiß, wem Maria wann was gegeben hat?
  who knows whom Maria when what given has

  b. $\lambda p.\exists x[p = \lambda w.[\mathsf{know'}_w(x, \lambda p.\exists y \exists z \exists t[\text{Maria gives } z \text{ to } y \text{ at } t])]]$

  c. $\lambda p.\exists x \exists z \exists t[p = \lambda w.[\mathsf{know'}_w(x, \lambda p.\exists y[\text{Maria gives } z \text{ to } y \text{ at } t])]]$

  d. \$ $\lambda p.\exists x \exists z[p = \lambda w.[\mathsf{know'}_w(x, \lambda p.\exists y \exists t[\text{Maria gives } z \text{ to } y \text{ at } t])]]$

  e. \$ $\lambda p.\exists x \exists t[p = \lambda w.[\mathsf{know'}_w(x, \lambda p.\exists y \exists z[\text{Maria gives } z \text{ to } y \text{ at } t])]]$

- Well-formedness constraints on underspecified representations are quite awkward to state such as the requirements

  - that the constraints should not lead to a cyclic plugging.
  - that every occurrence of a variable is properly bound.

- Depending on the semantic analysis, the two readings of (14) cannot be given a single underspecified representation.

  (14) Mary observed a man with a telescope.

- In Frank and Reyle (1995) and Egg (1998), the scope of a quantifier is underspecified but not its "reading", i.e., whether it is used distributively or collecitively etc.

  (15) Wer hat jedem bei der Vorbereitung geholfen? (Pafel, 1998)
  who has everyone with the preparation helped
  'Who helped everyone with the preparation?'

  a. Detmar.

  b. Frank helped Manfred; Janina helped Frank; Manfred helped Sabine; ....

- It is unclear for which phenomena underspecification should be chosen.

2.5 Summary

Systems with underspecified denotation raise conceptual questions about the limits of grammar and have severe technical limitations.

3 Indirect Representation: Ty2U^P

3.1 A Sketch

We can extend the appropriateness conditions of the sort *ur* to contain also a **plugging**.

(16) a. *undersp.-repr*

HOLES	*list(hole)*
LABELS	*list(label)*
CONSTR(AINTS)	*list(constraint)*
PLUGGING	*list(plug)*
MAIN	*label*
TOP	*hole*

b. *plug*

HOLE	*hole*
LABEL	*label*

In Ty2U^P the CONTENT value of a sentence contains a bijection between the elements in the HOLES and the LABELS list as one of its members.
Description of the CONTENT values:

(17) a. The ∀∃-reading:

$$\begin{bmatrix} \text{HOLES} & \langle h_0, h_1, h_2, h_3, h_4 \rangle \\ \text{LABELS} & \langle \boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5} \rangle \\ \text{CONSTR} & \langle \boxed{1} \leq h_0, \boxed{2} \leq h_0, \boxed{4} \leq h_0, \boxed{2} \leq h_1, \boxed{1} \leq h_2, \boxed{3} \leq h_0, \boxed{5} \leq h_0, \boxed{3} \leq h_3, \boxed{1} \leq h_4 \rangle \\ \text{PLUG} & \langle h_0 = \boxed{5}, h_1 = \boxed{2}, h_2 = \boxed{1}, h_3 = \boxed{3}, h_4 = \boxed{4} \rangle \\ \text{MAIN} & \boxed{1} \\ \text{TOP} & h_0 \end{bmatrix}$$

b. The ∃∀-reading:

$$\begin{bmatrix} \text{HOLES} & \langle h_0, h_1, h_2, h_3, h_4 \rangle \\ \text{LABELS} & \langle \boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5} \rangle \\ \text{CONSTR} & \langle \boxed{1} \leq h_0, \boxed{2} \leq h_0, \boxed{4} \leq h_0, \boxed{2} \leq h_1, \boxed{1} \leq h_2, \boxed{3} \leq h_0, \boxed{5} \leq h_0, \boxed{3} \leq h_3, \boxed{1} \leq h_4 \rangle \\ \text{PLUG} & \langle h_0 = \boxed{3}, h_1 = \boxed{2}, h_2 = \boxed{5}, h_3 = \boxed{4}, h_4 = \boxed{1} \rangle \\ \text{MAIN} & \boxed{1} \\ \text{TOP} & h_0 \end{bmatrix}$$

3.2 Summary

- The empirical problems of Ty2U disappear.
- The technical problems remain.
- The effect of having a single semantic representation for ambiguous sentences is gone.

4 Discontinuous Representation: LRS (Richter and Sailer, 2000)

4.1 A Sketch

(18) The sort *lrs*

$$\begin{array}{lll} \textit{lrs} & \textsc{top} & \textit{me} \\ & \textsc{main} & \textit{me} \\ & \textsc{parts} & \textit{list(me)} \end{array}$$

(19) Every student read some book

Relevant parts of the lexical entries:

(20) a. *read*:
$$\begin{bmatrix} \textit{lrs} & \\ \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{1}\ (\mathsf{read}'y)x \\ \textsc{parts} & \left\langle x, y, \boxed{1}, \boxed{1a}\mathsf{read}'y, \boxed{1b}\mathsf{read}' \right\rangle \end{bmatrix}$$

b. *book* :
$$\begin{bmatrix} \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{2}\mathsf{book}'(y) \\ \textsc{parts} & \left\langle y, \boxed{2}, \boxed{2a}\mathsf{book}' \right\rangle \end{bmatrix}$$

c. *student*:
$$\begin{bmatrix} \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{3}\mathsf{student}'(x) \\ \textsc{parts} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{student}' \right\rangle \end{bmatrix}$$

d. *some*:
$$\begin{bmatrix} \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{4}\exists y[\boxed{a} \wedge \boxed{b}] \\ \textsc{parts} & \left\langle y, \boxed{4}, \boxed{4a}\ [\boxed{a} \wedge \boxed{b}] \right\rangle \end{bmatrix}$$

e. *every*:
$$\begin{bmatrix} \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{5}\forall x[\boxed{c} \rightarrow \boxed{d}] \\ \textsc{parts} & \left\langle x, \boxed{5}, \boxed{5a}\ [\boxed{c} \rightarrow \boxed{d}] \right\rangle \end{bmatrix}$$

(21)
$$\begin{bmatrix} \textit{lrs} & \\ \textsc{top} & \textit{me} \\ \textsc{main} & \boxed{5} \begin{bmatrix} \textit{universal} & \\ \textsc{var} & x \\ \textsc{scope} & \boxed{5a} \begin{bmatrix} \textit{implication} & \\ \textsc{arg1} & \boxed{c}\ \textit{me} \\ \textsc{arg2} & \boxed{d}\ \textit{me} \end{bmatrix} \end{bmatrix} \\ \textsc{parts} & \left\langle \boxed{5}, \boxed{5a}, x \right\rangle \end{bmatrix}$$

Combinatorial Semantics:
We impose two wellformedness conditions on *lrs*:

(22) The MAIN PRINCIPLE (MP):

In each *lrs*, the MAIN value is an element of the PARTS list and a component of the TOP value.

(23) The TOP PRINCIPLE (TP):

1. In every phrase, the TOP value of the non-head daughter is an element of its PARTS list.
2. In every clause, the TOP value is the largest term of the PARTS list that consists exactly of the MAIN value's components and elements of the PARTS list.
3. In every utterance, every subterm of the TOP value of the utterance is an element of its PARTS list, and every element of the utterance's PARTS list is a subterm of the TOP value.

(24) SEMANTICS PRINCIPLE

In each *headed-phrase*,

1. the TOP value of the head and the mother are identical,
2. the MAIN value of the head and the mother are identical,[1]
3. the PARTS value contains exactly all elements of the PARTS values of the daughters,
4. the following conditions hold:
   (a) if the nonhead is a quantifier then its TOP value is of the form $Qx[\rho \circ \nu]$[2], the MAIN value of the head is a component of ρ, and the TOP values of both daughters are identical,
   (b) if the non-head is a quantified NP with a TOP value of the form $Qx[\rho \circ \nu]$, then the MAIN value of the head is a component of ν,
   (c) in a *head-adjunct-phrase*, the TOP value of the non-head is a component of the TOP value of the head, and
      i. if the non-head is an intersective modifier, then its TOP value is of the form $\alpha \wedge \beta$ and the MAIN value of the head is a component of β.
      ii. if the nonhead is a non-intersective modifier, then it is of the form $\alpha(\beta)$ and the MAIN value of the head is a component of β.

4.2 A Case Study: Negative Concord in Polish

4.2.1 Data

N-Words are inherently negative:

(25) a. Kogo widziałeś? Nikogo.
Who have you seen? Nobody.GEN/ACC.

b. Ile przeczytałeś książek?
How many books have you read?

Żadnej./ Żadną.
None.GEN./ None.ACC.

(26) Chcę poślubić albo Piotra, albo nikogo.
I will marry either Piotr or nobody

(27) Kocham ją jak [żadną inną].
I love her.ACC as [no other].ACC

? Kocham ją jak [żadnej innej].
I love her.ACC as [no other].GEN

'I love her more than (I love) any other (girl).'

Non-negative n-words?

(28) Janek *(nie) pomaga nikomu.
Janek NM helped nobody
'Janek didn't help anybody.'
$ 'Janek didn't help nobody.'

[1] We take the noun to be the head of a quantified NP.

[2] $Qx[\rho \circ \nu]$ is shorthand for the description $\begin{bmatrix} \textit{quantifier} & \\ \text{VAR} & \textit{var} \\ \text{SCOPE} & \begin{bmatrix} \textit{l-const} & \\ \text{ARG1} & \rho \\ \text{ARG2} & \nu \end{bmatrix} \end{bmatrix}$.

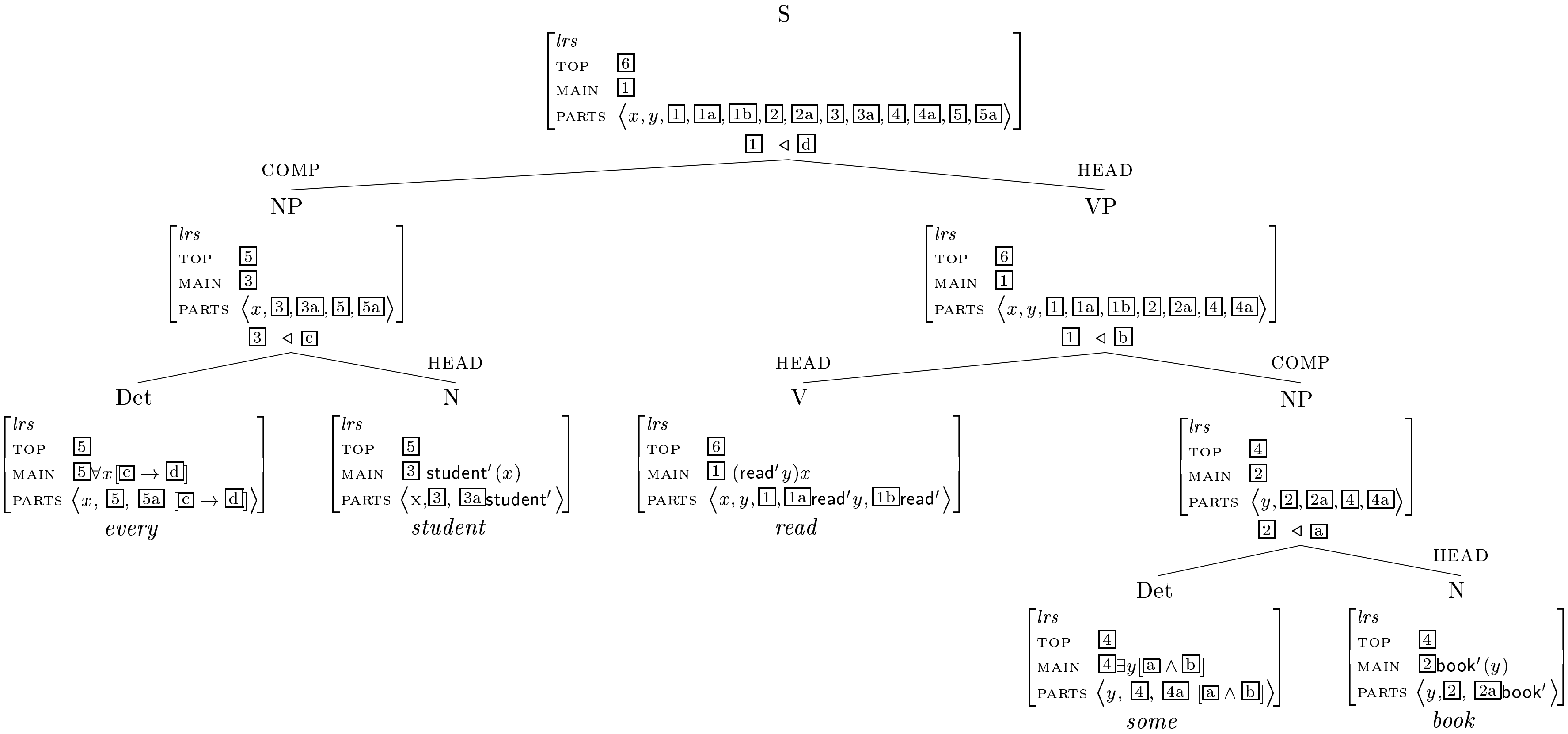


Figure 2: The LRS analysis of the sentence *Every student read some book.*:

(29) Nikt *(nie) pomaga nikomu.
Nobody NM helped nobody
'Nobody helped anybody.'

(30) Zaczął bez czekania na nikogo.
he-started without waiting for nobody
'He started without waiting for anybody.'

(31) * Widziałeś nikogo?
you-saw nobody
$ 'Did you see anybody?'

. . . but!

1. All these examples are interpreted as negative.
2. Polish n-words are not NPIs since they do not occur in the same envirionments as NPIs (only negativ contexts).

Conclusion There EXISTS evidence for the inherently negative character of Polish n-words.
There is NO evidence for an inherently non-negative character.

4.2.2 Analysis

Description of the word *nie przyszedł* (*NM came*):

(32)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nie przyszedł} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{TOP} & \textit{me} \\ \text{MAIN} & \boxed{1}\ \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}', \boxed{2}\neg\boxed{a}, \right\rangle \end{bmatrix} \end{bmatrix}$$
$\&\ \boxed{1} \triangleleft \boxed{a}$

Relevant parts of the lexical entry of *nikt* (*nobody*):

(33)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nikt} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{TOP} & \boxed{4a}\ \exists x \boxed{b} \\ \text{MAIN} & \boxed{3}\ \mathsf{human}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{human}', \boxed{4}\neg\exists x\boxed{b}, \boxed{4a}\exists x\boxed{b}, \boxed{5}[\boxed{c} \wedge \boxed{d}] \right\rangle \end{bmatrix} \end{bmatrix}$$
$\&\ \boxed{5} \triangleleft \boxed{b}$
$\&\ \boxed{3} \triangleleft \boxed{c}$

The analysis of *Nikt nie przyszedł* (*Nobody came*):

(34)
$$\begin{bmatrix} \text{TOP} & \boxed{6} \\ \text{MAIN} & \boxed{1} \\ \text{PARTS} & \left\langle x, x, \boxed{1}, \boxed{1a}, \boxed{2}, \boxed{3}, \boxed{3a}, \boxed{4}, \boxed{4a}, \boxed{5} \right\rangle \end{bmatrix}$$

Nikt
$$\begin{bmatrix} \text{TOP} & \boxed{4a} \\ \text{MAIN} & \boxed{3} \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}, \boxed{4}, \boxed{4a}, \boxed{5} \right\rangle \end{bmatrix}$$

nie przyszedł
$$\begin{bmatrix} \text{TOP} & \boxed{6} \\ \text{MAIN} & \boxed{1} \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}, \boxed{2} \right\rangle \end{bmatrix}$$

$\&\ \boxed{1} \triangleleft \boxed{a}$
$\&\ \boxed{5} \triangleleft \boxed{b}$
$\&\ \boxed{3} \triangleleft \boxed{c}$
$\&\ \boxed{1} \triangleleft \boxed{d}$

The following TOP values are possible:

(35)
a. $\neg\neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{4} \triangleleft \boxed{2} = \boxed{6})$ $= \exists x[\mathsf{come}'(x)]$
b. $\neg\exists x\neg[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{2} \triangleleft \boxed{4} = \boxed{6})$ $= \forall x[\mathsf{come}'(x)]$
c. $\neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{2} = \boxed{4} = \boxed{6})$

As only the third reading is available in Polish, we introduce the following, language-specific constraint:

The Negation Complexity Constraint:

(36) The NEGATION COMPLEXITY CONSTRAINT

For each sign, there may be at most one negation that is a component of the TOP value and has the MAIN value as its component.

(37) Poślubię albo tę dziewczynę z Poznania, albo żadnej dziewczyny z żadnego miasta.
I will marry either [this girl].ACC from Poznan or [no girl].GEN from no city.

'I will either marry this girl from Poznan or no girl from any city.'

(38) Tomek nie może nie znać Marii.
Tomek NM may NM know Maria

'It is not the case that it is possible that Tomek does not know Maria.'

The Neg Criterion:

(39) The NEG CRITERION:

For every verb, if there is a negation in the TOP value of the verb that has scope over the verb's MAIN value, then that negation must be an element of the verb's PARTS list.

(40) a. * Nikt przyszedł.
Nobody came

b. Relevant parts of the lexical entry of *przyszedł* (*came*):

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{przyszedł} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{TOP} & \textit{me} \\ \text{MAIN} & \boxed{1}\mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}' \right\rangle \end{bmatrix} \end{bmatrix}$$

c.

$$\begin{bmatrix} \text{TOP} & \boxed{4}\ \neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)] \\ \text{MAIN} & \boxed{1}\ \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}, \boxed{3}, \boxed{3a}, \boxed{4}, \boxed{4a}, \boxed{5} \right\rangle \end{bmatrix} \quad \begin{array}{l} \& \ \boxed{5} \triangleleft \boxed{b} \\ \& \ \boxed{3} \triangleleft \boxed{c} \\ \& \ \boxed{1} \triangleleft \boxed{d} \end{array}$$

Nikt

$$\begin{bmatrix} \text{TOP} & \boxed{4a} \\ \text{MAIN} & \boxed{3}\ \mathsf{human}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{human}', \boxed{4}\neg\exists x\boxed{b}, \boxed{4a}\exists x\boxed{b}, \boxed{5}[\boxed{c} \wedge \boxed{d}] \right\rangle \end{bmatrix}$$

przyszedł

$$\begin{bmatrix} \text{TOP} & \boxed{4} \\ \text{MAIN} & \boxed{1}\ \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}' \right\rangle \end{bmatrix}$$

4.2.3 Evaluation of the analysis

- Our analysis avoids mechanisms such as *negative absorption*:
  $\forall x_1 \neg\forall x_2 \ldots \neg\forall x_n \neg\phi \Longrightarrow \forall x_1 \forall x_2 \ldots \forall x_n \neg\phi$
- instead, it uses *identities*:
  - identities are the major analytical device of HPSG
  - identities arise wherever possible, so NC languages are predicted to be the unmarked case.
- The Negation Complexity Constraint and the Neg Criterion are adaptations of standard principles.

4.3 Summary

- LRS does not introduce an additional level of "underspecification" but makes use of the fact that
  - a sign contains its logical form as its component
  - one description can characterize many distinct objects
  - as these objects are terms, we have a natural place for "underspecification" in the grammar,
  - but this does not lead to an underspecification in the denotation nor to an indirect representation.
- the empirical and the technical problems of denotationally underspecified or indirect systems are avoided.

5 Fully Specified Denotation, Direct and Continuous Representation: LF-Ty2 (Richter and Sailer, 1999a,b; Sailer, 2000)

5.1 A Sketch

LF-Ty2 is based on the flexible type shifting system of Hendriks (1993) (see also Dekker (1993), Bouma (1994)):

- there is a basic translation for every word
- type shifting rules can freely apply to the translation of words
- the logical form of a phrase is the (intensional) functional application of the logical forms of its daughters.

Basic translations:

(41) a. *read* $\rightsquigarrow \lambda y \lambda x.\mathsf{read}'(x,y)$

b. *book* $\rightsquigarrow \lambda y.\mathsf{book}'(y)$

c. *student* $\rightsquigarrow \lambda y.\mathsf{student}'(y)$

d. *some* $\rightsquigarrow \lambda P \lambda Q.\exists y[P(x) \wedge Q(x)]$

e. *every* $\rightsquigarrow \lambda P \lambda Q.\forall y[P(y) \rightarrow Q(y)]$

Type shifting rule:

(42) Argument Raising:
$\lambda x_1 \ldots \lambda x_i \ldots \lambda x_n.\phi \longrightarrow_{AR_i} \lambda x_1 \ldots \lambda X_i \ldots \lambda x_n.X_i(\lambda x_i.\phi)$

5.2 Comparison to LRS

- LF-Ty2 is closer to standard techniques in semantics.
- but the discontinuous representation in LRS allows us
  - to avoid complicated mechanism for combinatorial semantics such as λ conversion or type shifting,
  - to use identities in combinatorial semantics.

Figure 3: The ∀∃ reading of the sentence *Every student read some book*:

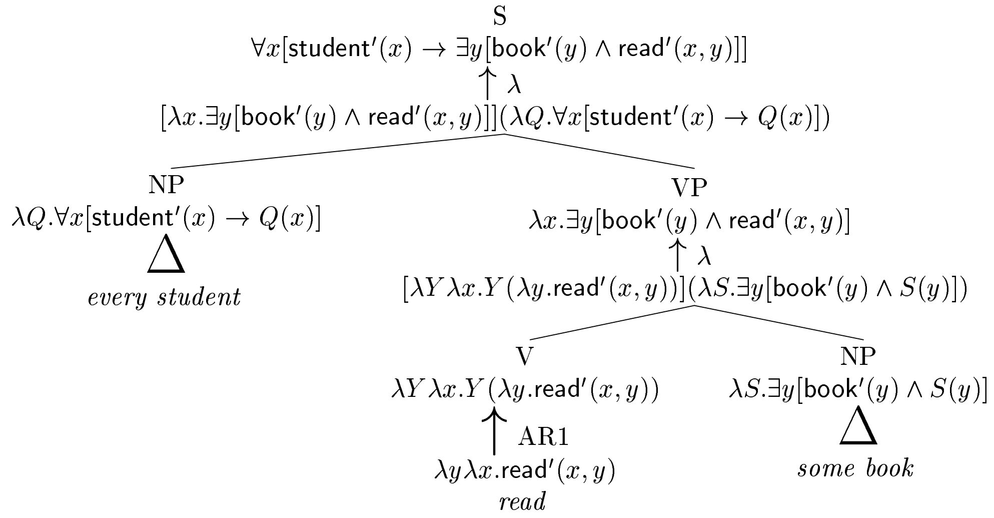


Figure 4: The ∃∀ reading of the sentence *Every student read some book*:

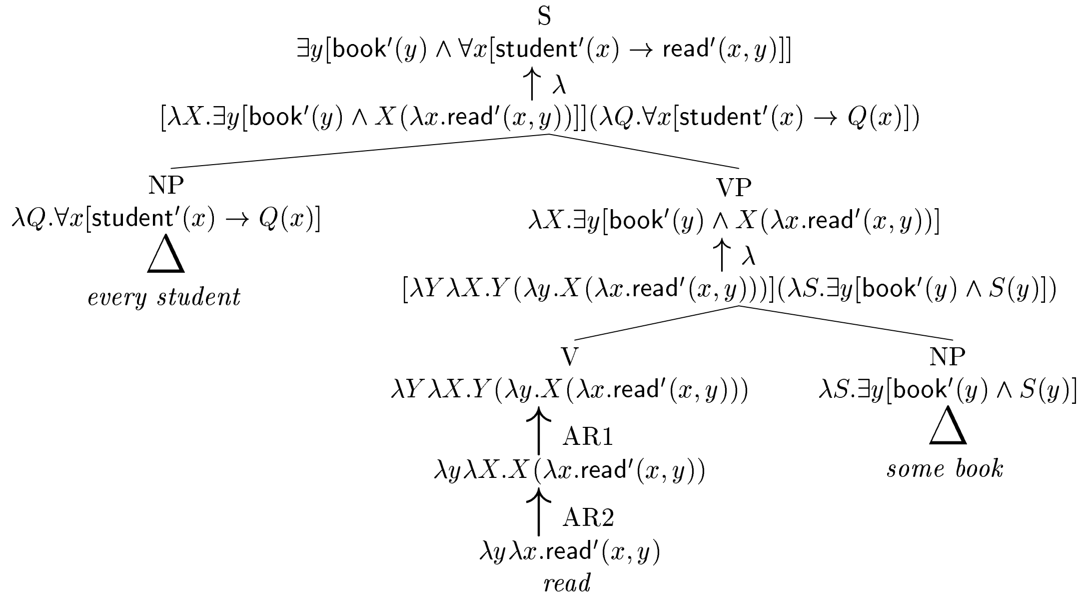


Table 2: Classification of semantic proposals for HPSG

Approach	underspecified den.	indirect repr.	discontinuous repr.
Ty2U	+	+	+
Ty2U^P	-	+	+
LRS	-	-	+
LF-Ty2	-	-	-
UDRT	+?	+	+
MRS	?	+	+
UMRS	+?	+	+

6 Discussion

6.1 Where are MRS, UMRS, and UDRT?

As can be seen in the table, most HPSG proposals for "underspecified semantics" are *indirect* systems, or somewhere between indirect and denotationally underspecified systems.

- UDRT (Frank and Reyle, 1995)
  - Scoping relations are left open.
  - An extra-grammatical resolution procedure is used to resolve the ambiguity between collective and distributive readings.
- MRS (Copestake et al., 1997))

  > "The MRS structures generated by a grammar will typically not themselves be scope-resolved MRS structures, but will correspond to a set of such structures. The condition on MRS structures in general is that it must be possible to further instantiate the structure to produce a scope-resolved MRS structure (without free variables) by adding coindexation statements between argument handles and (one or more) labels: i.e., by adding outscopes relations and (possibly) equating labels. Instantiation may be partial, that is it need not result in a fully scope-resolved MRS." (Copestake et al., 1997)

  Main differences between MRS and the approaches presented in this talk:
  - unclear whether it is a denotationally underspecified or an indirect system.
  - Constructional aspects: phrases can introduce idiosyncratic semantic contributions.
  - There is only a loose connection between the objects denoted by the grammar and terms of a semantic representation language.
- UMRS (Egg and Lebeth, 1995; Egg, 1998)
  - An extra-grammatical *translation procedure* is given from UMRS structures to terms of some semantic representation language. Only some UMRS structures can be the input to this procedure, in particular they must be acyclic.

    > "This rule prevents ill-formed disambiguation of UMRS structures (which would be too difficult to filter out in the metalanguage itself) ..." (Egg, 1998, p. 46)
  - Some disambiguation is encoded in the sort hierarchy, in particular the distinction between a normal universal quantifier and a universal quantifier which outscopes an interrogative quantifier, as assumed for the pair-list reading in (15).

6.2 Open Questions

- Is there a similar space of possibilities in TAG?
- Why did Kallmeyer and Joshi (2000) choose a denotationally underspecified system?
- Why are indirect systems attractive?
- How big is the advantage of discontinuous systems?
- Can we decide between these systems on the basis of data?

References

Beck, Sigrid (1996). *Wh-Constructions and Transparent Logical Form.* Ph. D. thesis, Universität Tübingen.

Bos, Johan (1995). Predicate Logic Unplugged. In *Proceedings of the 10th Amsterdam Colloquium.*

Bouma, Gosse (1994). Calculated Flexibility. In H. Bunt, R. Muskens, and G. Rentier (Eds.), *Proceedings of the International Workshop on Computational Semantics*, pp. 32–40. Katholieke Universiteit Brabant.

Copestake, Flickinger, Dan, and Sag, Ivan (1997). Minimal Recursion Semantics. An Introduction. Manuscript, Stanford University, URL: `ftp://ftp-csli.stanford.edu/linguistics/sag/mrs.ps.gz`.

Dekker, Paul (1993). *Ups and Downs in Dynamic Semantics.* Ph. D. thesis, Universiteit van Amsterdam.

Egg, Markus (1998). *Wh*-questions in Underspecified Minimal Recursion Semantics. *Journal of Semantics 15*, 37–82.

Egg, Markus and Lebeth, Kai (1995). Semantic underspecification and modifier attachment. In *Proceedings der Jahrestagung der DGfS.*

Frank, Anette and Reyle, Uwe (1995). Principle Based Semantics for HPSG. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, pp. 9–16. Association for Computational Linguistics.

Gallin, Daniel (1975). *Intensional and Higher-Order Modal Logic.* North-Holland, Amsterdam.

Hendriks, Herman (1993). *Studied Flexibility.* ILLC Dissertation Series 1995-5. Institute for Logic, Language and Computation, Amsterdam.

Kallmeyer, Laura and Joshi, Aravind K. (2000). Fatoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG. to appear in *Language and Computation.*

King, Paul J. (1999). Towards Truth in HPSG. In V. Kordoni (Ed.), *Tübingen Studies in HPSG*, Arbeitspapiere des SFB 340, Nr. 132, Volume 2, pp. 301–352. Universität Tübingen.

Montague, Richard (1974). The Proper Treatment of Quantification in Ordinary English. In R. H. Thomason (Ed.), *Formal Philosophy. Selected Papers of Richard Montague*, pp. 247–270. Yale University Press.

Pafel, Jürgen (1998). Skopus und logische Struktur. Studien zum Quantorenskopus im Deutschen. Arbeitspapiere des SFB 340, Nr. 129, Universität Tübingen.

Pollard, Carl and Sag, Ivan A. (1994). *Head-Driven Phrase Structure Grammar.* University of Chicago Press.

Pollard, Carl Jesse (1999). Strong Generative Capacity in HPSG. In G. Webelhuth, J.-P. Koenig, and A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Chapter 18, pp. 281–297. CSLI Publications.

Przepiórkowski, Adam and Kupść, Anna (1999). Eventuality Negation and negative Concord in Polish and Italian. In R. D. Borsley and A. Przepiórkowski (Eds.), *Slavic in HPSG*, pp. 211–246. CSLI Publications.

Richter, Frank (2000). A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar. Dissertation, Universität Tübingen, Version of April 28th, 2000.

Richter, Frank and Sailer, Manfred (1999a). A Lexicalist Collocation Analysis of Sentential Negation and Negative Concord in French. In V. Kordoni (Ed.), *Tübingen Studies in Head-Driven Phrase Structure Grammar*, Arbeitspapiere des SFB 340, Nr. 132, Volume 1, pp. 231–300. Universität Tübingen.

Richter, Frank and Sailer, Manfred (1999b). LF Conditions on Expressions of Ty2: An HPSG Analysis of Negative Concord in Polish. In R. D. Borsley and A. Przepiórkowski (Eds.), *Slavic in HPSG*, pp. 247–282. CSLI Publications.

Richter, Frank and Sailer, Manfred (1999c). Underspecified Semantics in HPSG. In H. Bunt and R. Muskens (Eds.), *Computing Meaning*, Studies in Linguistics and Philosophy, pp. 95–112. Kluwer academic publishers.

Richter, Frank and Sailer, Manfred (2000). On the Left Periphery of German Finite Sentences. In D. Meurers and T. Kiss (Eds.), *Topics in Constraint-Based Approaches to Germanic Syntax (working title)*. CSLI Publications. Forthcoming.

Richter, Frank, Sailer, Manfred, and Penn, Gerald (1999). A Formal Interpretation of Relations and Quantification in HPSG. In G. Bouma, E. Hinrichs, G.-J. M. Kruijff, and R. T. Oehrle (Eds.), *Constraints and Resources in Natural Language Syntax and Semantics*, pp. 281–298. CSLI Publications.

Sailer, Manfred (2000). Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar. Dissertation, Eberhard-Karls-Universität Tübingen, Version of June 29th, 2000.