Basic Concepts of Lexical Resource Semantics*

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

June 30, 2003

1 Introduction

Semanticists use a range of highly expressive logical languages to characterize the meaning of natural language expressions. The logical languages are usually taken from an inventory of standard mathematical systems, with which generative linguists are familiar. They are, thus, easily accessible beyond the borders of a given framework such as Categorial Grammar, Lexical Functional Grammar, or Government and Binding Theory. Linguists working in the HPSG framework, on the other hand, often use rather idiosyncratic and specialized semantic representations. Their choice is sometimes motivated by computational applications in parsing, generation, or machine translation. Naturally, the intended areas of application influence the design of semantic representations. A typical property of semantic representations in HPSG that is concerned with computational applications is underspecification, and other properties come from the particular unification or constraint solving algorithms that are used for processing grammars. While the resulting semantic representations have properties that are motivated by, and are adequate for, certain practical applications, their relationship to standard languages is sometimes left on an intuitive level. In addition, the theoretical and ontological status of the semantic representations is often neglected. This vagueness tends to be unsatisfying to many semanticists, and the idiosyncratic shape of the semantic representations confines their usage to HPSG. Since their entire architecture is highly dependent on HPSG, hardly anyone working outside of that framework is interested in studying them.

With our work on Lexical Resource Semantics (LRS), we want to contribute to the investigation of a number of important theoretical issues surrounding semantic representations and possible ways of underspecification. While LRS is formulated in a constraint-based grammar environment and takes advantage of the tight connection between syntax proper and logical representations that can easily be achieved in HPSG, the architecture of LRS remains independent from that framework, and combines attractive properties of various semantic systems. We will explore the types of semantic frameworks which can be specified in Relational Speciate Re-entrant Language (RSRL), the formalism that we choose to express our grammar principles, and we evaluate the semantic frameworks with respect to their potential for providing empirically satisfactory analyses of typical problems in the semantics of natural languages. In LRS, we want to synthesize a flexible meta-theory that can be applied to different interesting semantic representation languages and make computing with them feasible.

We will start our investigation with a standard semantic representation language from natural language semantics, Ty2 (Gallin, 1975). We are well aware of the debate about the appropriateness of Montagovian-style intensionality for the analysis of natural language semantics, but we believe that it is best to start with a semantic representation that most generative linguists are familiar with. As will become clear in the course of our discussion, the LRS framework is a meta-theory of semantic representations, and we believe that it is suitable for various representation languages,

*This paper can be regarded as a snapshot of our work on LRS. It was written as material for the authors' course *Constraint-based Combinatorial Semantics* at the 15th European Summer School in Logic, Language and Information in Vienna in August 2003. It is meant as background reading and as a basis of discussion for our class. Its air of a *work in progress* is deliberate. As we see continued development in LRS, its application to a wider range of languages and empirical phenomena, and especially the implementation of an LRS module as a component of the TRALE grammar development environment; we expect further modifications and refinements to the theory. The implementation of LRS is realized in collaboration with Gerald Penn of the University of Toronto. We would like to thank Carmella Payne for proofreading various versions of this paper.

including alternative, non-intensional systems. In most of the introductory examples of this paper, we will ignore intensionality. To give a concrete example of our intentions, (1) shows the two semantic representations, (1b) and (1c), that will be assigned to the sentence in (1a).

(1) a. Everyone reads something.

b. $\forall x \exists y[\mathsf{read}'(x, y)]$

c. $\exists y \forall x[\mathsf{read}'(x, y)]$

We will precede as follows: After a brief introduction to the formalism, the architecture of grammar, and the encoding of the expressions of Ty2 in RSRL, we will investigate various techniques of underspecification and of combining (information about) semantic representations of syntactic daughters at phrases.

The four semantic systems that we obtain by varying the options of underspecification and combination serve as a small taxonomy to classify the current semantic representation languages of the HPSG literature. We distinguish systems with a truly underspecified denotation; systems that use an indirect representation; systems with discontinuous representations; and systems that only employ classical, "fully specified" representations with a traditional combinatorial system such as lambda abstraction and functional application. The usual model-theoretic interpretation of the expressions of Ty2 that underly all systems will allow us to observe the behavior and consequences of the modifications with respect to theoretical and empirical linguistic issues while keeping the intended interpretation of the natural language expressions constant. Drawing on well-known empirical phenomena in various languages, we will show that different choices force the linguist to express certain empirical generalizations in different modules of the grammar. Depending on the module in which the generalizations are expressed, different natural classes of data are predicted. To the extent to which these predictions can be tested, the choice between the semantic systems becomes accessible to empirical tests. Besides arguments of computational feasibility and largely aesthetical differences in elegance, we obtain an additional dimension for choosing a suitable semantic system for constraint-based grammars of natural languages.

On the basis of our observations about the properties of the different architectures for semantic representations, we propose LRS as an adequate meta-theory of semantic representation. LRS combines common assumptions about semantic representations with techniques of linguistic description in constraint-based grammar. Our main goal throughout this study is a theoretically satisfying, as well as empirically and computationally viable specification of the meaning of natural language expressions. We hope that one of the benefits of our current project will be to increase interest in future work on the syntax-semantics interface of HPSG.

2 Foundations

2.1 The Formal and Linguistic Framework

2.1.1 RSRL

We choose *Relational Speciate Re-entrant Language* (RSRL, Richter et al. (1999); Richter (2000)) as the description language for HPSG grammars. In RSRL, a grammar consists of a *signature* and a *theory*.[1] The signature consists of a sort hierarchy, a set of attributes, a set of relation symbols together with an arity function that assigns each relation symbol a positive natural number, and appropriateness conditions, which are sometimes called *feature declarations*. Appropriateness determines which attributes are appropriate to which sort, and which values are appropriate for an attribute at a given sort. For example, it is standardly assumed that SYNSEM is an attribute that is appropriate for *sign*, and the values of the SYNSEM attribute at *signs* are of sort *synsem*. The theory is a set of descriptions, the *principles* of grammar. In traditional HPSG grammars, the theory contains principles such as the HEAD FEATURE PRINCIPLE, the SUBCATEGORIZATION PRINCIPLE, the ID PRINCIPLE, the WORD PRINCIPLE, the SEMANTICS PRINCIPLE, and the principles of the BINDING THEORY, among many others.

A *model* of a grammar is a collection of objects whose components are configured in accordance with the signature; and all objects, including all their components, satisfy each description in the theory. Typically, these objects are

[1]This assumption is consistent with Pollard and Sag (1994); King (1999); Pollard (1999).

constructed as totally well-typed, sort resolved, relational abstract feature structures; but other possibilities have been suggested as well. Essentially, there are three theories of the meaning of HPSG grammars available, which differ with respect to their philosophical assumptions about the nature of linguistic theories.

According to King (1999), an *exhaustive model* of a grammar contains “instances” of all objects that are licensed by the theory. One model of the class of exhaustive models of a grammar contains the *possible tokens* of the natural language under consideration.

Alternatively, a collection of structures that can be construed in various ways is understood as the *types* of the natural language under consideration. The types are taken to be (a certain kind of) feature structures in (Pollard and Sag, 1994). Pollard and Sag regard these objects as belonging to the real world. Pollard (1999) retains types, but takes a slightly different (more agnostic) ontological perspective. In his view, the elements of the *strong generative capacity* of a grammar are mathematical idealizations of natural language expressions. Pollard defines the strong generative capacity of a grammar by combining King’s denotation functions with a variant of techniques known from constructing feature structure models.

The three different views of the meaning of HPSG grammars are discussed in Richter (2000). For our present purposes, there are no significant differences between them.

2.1.2 The Structure of Signs in HPSG

Expressed within the signature of the grammar, we find universal assumptions about the structure of linguistic signs. The structure of signs in HPSG is best visualized with a sketch describing all possible signs:

(2)
$$\begin{bmatrix} \textit{sign} & \\ \text{PHONOLOGY} & \textbf{phonological structure} \\ \text{SYNSEM} & \begin{bmatrix}\text{LOCAL} & \begin{bmatrix}\text{CATEGORY} & \textbf{(local) syntactic structure}\\ \text{CONTENT} & \textbf{semantic structure}\end{bmatrix}\end{bmatrix} \\ \text{DAUGHTERS} & \textbf{syntactic structure}\end{bmatrix}$$

Any expression of a natural language comprises at least phonological, syntactic and semantic structure. A crucial property of this architecture is the fact that for any given pair of phonological and syntactic structure, there might be various possible semantic structures in the denotation of a grammar. If a sentence is ambiguous, and the ambiguity has no syntactic counterpart, we will find as many non-isomorphic structures of that sentence in the denotation of the grammar as it has readings. For example, if we assume that there is no underspecification on the level of semantic structures, the sentence with the phonology given in (1a) has two non-isomorphic structures. Their syntactic and phonological part are isomorphic, but their semantic structures differ: (1b) is the semantic structure of one version of the sentence, (1c) the semantic structure of the other.

2.1.3 Underspecification

The most natural notion of *underspecification* in HPSG is *underspecification at the description level*. A description might denote many well-formed expressions. Given a realistic grammar of English, the description $\begin{bmatrix}\textit{phrase}\end{bmatrix}$ denotes an infinite amount of English structures, since there is an infinite amount of English expressions. Grammars may employ descriptive underspecification in the lexicon or in grammar principles. Let us first consider *underspecification in the lexicon*.

A lexical entry is a description. We may say that a lexical entry is underspecified if there is some attribute whose value is not uniquely fixed 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*. Typical semantic instances of that kind of underspecification in the lexicon of Pollard and Sag (1994) are “CONTENT raisers” like the verbs *to* and *be*. They share their CONTENT value with the verbal projection they select. In effect, they have infinitely many possible CONTENT values, depending on the syntactic context in which they occur.

The second possibility of descriptive underspecification is underspecification of grammatical structures by grammatical principles. In the grammar of Pollard and Sag (1994) sentence (1a) 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, their SEMANTICS PRINCIPLE is an instance of descriptive underspecification.

2.2 Assumptions about Semantic Representations

With our assumptions about the semantic structures of natural language expressions, we depart radically from (Pollard and Sag, 1994). Instead of postulating structures that are specific for HPSG, we postulate semantic structures that stand in one-to-one correspondence to a logical language. This gives us an opportunity to apply various techniques of underspecification that were developed for logical languages, particularly in the literature of computational semantics. What is new in our investigation compared to previous work is that we can test the predictions of each variant from our semantic representations while keeping the rest of the grammar unchanged. Nothing but the semantic representations change. For this reason, we can observe the changing predictions of the grammar in the models of the grammar as a function of different techniques of underspecification.

2.2.1 Expressions of Ty2

We assume that the CONTENT value of a sign is a *logical form*. In particular, we adopt the language Ty2 of Gallin (1975), which is similar to Intensional Logic (Montague, 1974) but has technical advantages. We also presuppose that we can specify a grammar of Ty2 in RSRL.

For an encoding of a grammar of Ty2 in RSRL, it is necessary to define an appropriate signature with sorts and attributes that provide the primitives of the expressions. The theory of Ty2 must ensure that in the denotation of the new sorts, we find all of the well-formed expressions of the target language and nothing else.[2]

The following example illustrates our encoding of Ty2 in RSRL:

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

$$\left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{abstraction}}\\
\textsc{type} & \boxed{0}\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{c-type}}\\ \textsc{in} & \boxed{1}\ \textit{entity}\\ \textsc{out} & \boxed{2}\ \textit{truth}\end{array}\right]\\
\textsc{var} & \boxed{3}\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{var}}\\ \textsc{type} & \boxed{1}\end{array}\right]\\
\textsc{arg} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{application}}\\
\textsc{func} & \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{student}}\\ \textsc{type} & \boxed{0}\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{c-type}}\\ \textsc{in} & \boxed{1}\\ \textsc{out} & \boxed{2}\end{array}\right]\end{array}\right]\\
\textsc{arg} & \boxed{3}
\end{array}\right]
\end{array}\right]$$

The AVM in (3) describes the expression $\lambda x_e.\mathsf{student}'_{et}(x_e)$. If we assume (with (Pollard and Sag, 1994)) that the models of our grammar contain abstract feature structures, then the AVM description denotes an abstract feature structure which we may interpret as a graph representation of the expression $\lambda x_e.\mathsf{student}'_{et}(x_e)$. Since we presuppose an RSRL specification of the expressions of Ty2, and since we know that for each expression of Ty2, we can write an AVM description that describes it (and no other expression), we henceforth simplify our notation and write expressions of Ty2 inside AVMs instead of their descriptions. This notational simplification enhances the readability of our grammatical descriptions significantly.

2.2.2 A Taxonomy

As we investigate the consequences of embedding expressions of Ty2 into grammars of natural languages, we vary three parameters of semantic systems. All three parameters have to do with different kinds of underspecification that can be applied to logical object languages such as first order predicate logic or some higher-order intensional language. They are usually introduced in the analysis of the meaning of natural languages in order to make a specification computationally tractable or to cope with phenomena where the composition of natural language expressions seems to behave markedly different from the simple composition rules that one would prefer to use with formal languages.

A typical application of underspecification for predominantly computational reasons concerns scope ambiguities. The syntactic form of sentences containing several quantifiers often does not determine the relative scope of the

[2]See Sailer (2000) for the details of the signature and an appropriate theory and the necessary correspondence proofs.

quantifiers. Even with a small number of quantifiers, resolving the possible readings in logical forms quickly leads to an explosion of the number of logical forms. For most practical purposes, however, complete scope resolution is unnecessary. Moreover linguistic observations indicate that humans do not entertain all possible readings when faced with those sentences either. However, the readings are available if needed.

The driving force behind our interest in underspecification is not computation, although we do keep an eye on practical considerations by developing an LRS module for TRALE. The main focus of our interest is a grammar theoretic understanding of what various treatments of object languages for logical forms imply for the kinds of predictions that we can make as grammar writers with principles at the syntax semantics interface.

The postulation of *discontinuous representations* is motivated by observations which suggest that the semantic contribution of words is not a single logical expression but a collection of expressions. Moreover, these expressions are usually discontinuously distributed over the logical form of the bigger syntactic units in which the words occur. In this view, the logical representations of the meanings of sentences cannot be divided into continuous units such that there is a bijection from the logical units to the lexical elements that introduce them.[3]

With the term *underspecified denotation* we are referring to the idea that the logical form, or, in HPSG's terminology, the CONTENT value of a sign, represents an underspecified semantic expression. This means that the grammar itself does not resolve the readings. The linguistic entities that are the subject of the study of language do not exhibit logical forms that stand for single readings. They may be resolved outside of the grammar, but that happens beyond the borders of linguistic description.

With the term *indirect representations* we are referring to the occurrence of meta-variables in logical forms (our CONTENT values). These meta-variables are variables with a denotation in the domain of the logical representation language, and they make it possible to leave room for plugging in other expressions.

Classical Montague semantics based on Intensional Logic (or variants thereof) do not use underspecified denotation, indirect representations, or discontinuous representations. Our system LF-Ty2 shows that a straightforward implementation in HPSG of Flexible Montague Grammar (Hendriks, 1993), a variant of the Montagovian machinery, is technically unproblematic.

Table 1: Classification of semantic representations in HPSG

Approach	underspecified denotation	indirect representation	discontinuous representation
Ty2U	+	+	+
Ty2U^P	-	+	+
LRS	-	-	+
LF-Ty2	-	-	-

Table 1 summarizes the possibilities that we are interested in. Each one of them is represented by a particular treatment of Ty2 as an object language. In the HPSG literature we find counterparts to them, which might employ radically different semantic structures, but whose capabilities of predicting data are subject to the same fundamental boundaries as our corresponding systems. Which kinds of data can be treated in the models of the grammars remains the same in all relevant respects.

In Section 3 we take a close look at the consequences of adopting the combination of properties represented by logical forms of type Ty2U, Ty2U^P, and LF-Ty2. In response to the insights gained in our investigation, we choose the combination of properties represented by LRS.

3 Existing Options for Semantic Representations

In this section we present encodings of systems that were originally developed outside of HPSG and outside of constraint-based grammar formalisms. The first of them, Ty2U, is in fact a specific instance or application of a more

[3] Assuming that only lexical units contribute pieces to the logical form. Alternatively, one could assume that the translation rules which combine the logical forms of the daughters at phrases might also contribute pieces to the overall semantic representation. This is of course the case in Montague's PTQ.

general system. The underlying mechanisms were originally conceived as a method to treat any traditional logical language and obtain an underspecified or "unplugged" version of it. This is essentially achieved by adding a layer to the language in which one talks about how fragments of an overall logical expression can be connected to form a specification of (potentially multiple) readings of a natural language expression. The system that we will examine in more detail, Ty2U, is an application of this general method to a specific logical language, taken from Ty2.

Problems that we observe for the description of certain empirical phenomena in different languages motivate a modification of Ty2U. The modification eliminates the property of allowing for underspecified denotations, because we identify this property as the source of the descriptive problems. The resulting system, Ty2U^P, however, loses the attractive features of the original system, whereas it retains all the technical difficulties originally associated with it. Eliminating the source of the problems this way does not seem to lead to satisfactory results.

The third system that we will review, LF-Ty2, is located at the other end of the spectrum. It associates each lexical entry with a continuous logical form, which may, through the application of certain techniques, systematically lead to more logical forms for the same lexical entry. In phrases, the logical forms of the syntactic daughters may only be combined by functional application. We will briefly indicate which technical machinery is needed to encode these basic ideas in HPSG. This will be sufficient to see that this architecture, while being satisfactory for formulating the necessary empirical generalizations over semantic phenomena, is wide open to the criticisms that originally led researchers to the development of underspecified systems in computational semantics.

3.1 Underspecified Denotation: Ty2U

The basic idea of Ty2U is that a sentence with multiple readings receives a unique syntactic and a unique semantic structure. In other words, the grammar denotes only one structure for ambiguous sentences. This, in turn, means that the semantic representation itself is vague, and its interpretation is a set of readings rather than a single reading. This approach to semantic representations as objects in the denotation of grammars allows the application of techniques described in the literature on underspecified semantic representation languages (c.f. Bos (1996), Pinkal (1996, 1999), Reyle (1993), among others). Furthermore, in a computational application, the number of HPSG structures to be computed for ambiguous sentences is reduced to one.

We give only a brief sketch of Ty2U. For more details see Richter and Sailer (1999c). The underspecification mechanism of Ty2U is directly taken from Bos (1996). Therefore, we will first present Bos's concepts and, then, show how these can be encoded as part of an HPSG grammar.

3.1.1 Bos (1996)

Bos (1996) defines an *underspecified representation* (UR) as a triple $\langle H, L, C\rangle$ that consists of a set of holes, a set of labeled formulae, and a set of constraints. The *holes* are meta-variables. A *labeled formula* is a labeled expression $l : \phi$ where l is a label and ϕ is an expression of some underspecified formal language. We extend the language of Gallin's Ty2 in such a way that its expressions may also contain meta-variables. In a labeled formula, the expression ϕ, then, is an expression of this extended Ty2. A *constraint* is an expression of the form $k \leq k'$, where k and k' are taken from the union of the set of holes and the set of labels. In our example, k always is a label and k' always is a hole. The expressions of the form $k \leq k'$ impose an order on the holes and labels. They will eventually be interpreted as requiring that any expression that is the interpretation of k be a subexpression of any expression that is the interpretation of k'. For clarity, we will sometimes call them *subordination constraints*.

In (4) the UR of the example sentence (1) is given explicitly. The first set introduces the meta-variables h_0, h_1 and h_2. The second provides the labels and the expressions. One expression is a regular Ty2 expression ($\mathsf{read}'(x,y)$), the other two contain a quantifier and a meta-variable ($\exists y[h_1]$ and $\forall x[h_2]$). The third set consists of subordination constraints.

(4) A set of holes: $\{h_0, h_1, h_2\}$
A set of labelled formulae: $\{l_1 : \mathsf{read}'(x,y),\quad l_2 : \exists y[h_1],\quad l_3 : \forall x[h_2]\}$
A set of constraints: $\{l_1 \leq h_0,\quad l_2 \leq h_0,\quad l_3 \leq h_0,\quad l_1 \leq h_1 \quad l_1 \leq h_2\}$

Such URs are usually depicted as graphs as in (5), where the nodes are either labeled formulae or holes; vertical domination in the graph indicates the constraints.

(5) A graphical representation of the UR in (4):

$$\begin{array}{ccc} & h_0 & \\ l_3 : \forall x.h_2 & & l_2 : \exists y.h_1 \\ & l_1 : \mathsf{read}'(x,y) & \end{array}$$

Bos (1996) is only interested in certain kinds of URs, namely URs that have a top element under $\leq$ with respect to a bijection from H to L. Such a bijection is called a *possible plugging*. The interpretation of a UR is the set of interpretations of the top-labeled formula of L under each possible plugging. In (6) the two possible pluggings are given for our example sentence. The two pluggings correspond to the two possible readings.

(6) a. $\forall x \exists y[\mathsf{read}'(x,y)]$: $P1 = \{h_0 = l_3, \quad h_1 = l_1, \quad h_2 = l_2\}$

b. $\exists y \forall x[\mathsf{read}'(x,y)]$: $P2 = \{h_0 = l_2, \quad h_1 = l_3, \quad h_2 = l_1\}$

When we interpret a UR under some possible plugging, we interpret the top hole with respect to the relation $\leq$. The interpretation of a hole is identical to that of the formula that bears the label that this hole is mapped to by the plugging. In the case of $P1$ in (6a), the hole h_0 is interpreted as the formula $\forall x[h_2]$, where the hole h_2 is interpreted as $\exists y.[h_1]$. Finally, the hole h_1 is interpreted as the expression $\mathsf{read}'(x,y)$. This yields the indicated $\forall\exists$-reading. In Bos (1996), URs can be interpreted directly, where the denotation of a UR is the union of the interpretations of all the possible pluggings of this UR.

3.1.2 A Sketch of Ty2U

In our HPSG encoding of Bos's URs we first assume that the CONTENT value of a sign is of some new sort, *underspecified-representation* (*ur*). The appropriateness conditions of this sort are given in (7). The values of the attributes HOLES, LABEL and CONSTR correspond to the three components of Bos's URs (a set of holes, a set of labeled formulae and a set of constraints). In addition to these three attributes, a *ur* object has two more attributes, MAIN and TOP, that are needed for the formulation of the SEMANTICS PRINCIPLE later in (10).

(7) *undersp.-repr*

	HOLES	*set(hole)*
	LABELS	*set(label)*
	CONSTR(AINTS)	*set(constraint)*
	MAIN	*label*
	TOP	*hole*

In (7) some new sorts are mentioned. We extend our sort hierarchy for encoding expressions of Ty2 to include objects of sort *hole* which will be interpreted as meta-variables. Furthermore, we introduce a sort *label*. Objects of this sort are interpreted as expressions of Ty2 which may contain meta-variables.[4] In addition, we postulate a supersort *meta* of the sorts *hole* and *label*. Finally we introduce a sort, *constr(aint)*, in order to capture the subordination constraints of Bos (1996). As indicated in (8), the attributes A1 and A2 are appropriate to *constr* objects and yield objects of sort *meta* as values.

(8) *constraint*

	A1	*meta*
	A2	*meta*

For readability, we write descriptions of *constraint* objects in the form $k \leq k'$, which we have seen before, where k is meant as the description of the A1 value and k' is the description of the A2 value. The symbol $\leq$ simply indicates in a convenient infix notation that we are talking about a *constraint* object.

[4]Since our formalism allows us to refer to logical expressions as individual entities, we do not have to introduce a mapping from labels to logical expressions in order to state subordination constraints elegantly. Instead, the "labelled" formulae themselves are used in the subordination constraints. There is no substantive difference between the two different ways of expressing the same idea.

With this signature, we can derive the correct UR of our example sentence. To this end, we need to specify the lexical entries of the words that occur in the sentence and to formulate a SEMANTICS PRINCIPLE that regulates the combinatorial semantics. In (9) the relevant parts of the lexical entries for *reads*, *something* and *everyone* are given.

(9) a. *read*:

$$\left[\begin{array}{ll} \text{PHON} & \langle \textit{reads} \rangle \\ \text{S L} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{SUBCAT} & \left\langle \text{NP, NP}\right\rangle\end{array}\right] \\ \text{CONT} & \left[\begin{array}{ll} \textit{ur} & \\ \text{HOLES} & \left\{h_0\right\} \\ \text{LABELS} & \left\{\boxed{1}\mathsf{read}'(x,y)\right\} \\ \text{CONSTR} & \left\{\boxed{1} \leq h_0\right\} \\ \text{MAIN} & \boxed{1} \\ \text{TOP} & h_0 \end{array}\right]\end{array}\right]\end{array}\right]$$

b. *something*:

$$\left[\begin{array}{ll} \text{PHON} & \langle \textit{something} \rangle \\ \text{S L} & \left[\begin{array}{ll} \text{CAT HEAD } \textit{noun} & \\ \text{CONT} & \left[\begin{array}{ll} \textit{ur} & \\ \text{HOLES} & \left\{h_0, h_1\right\} \\ \text{LABELS} & \left\{\boxed{2}\exists y[h_1]\right\} \\ \text{CONSTR} & \left\{\boxed{2} \leq h_0\right\} \\ \text{MAIN} & \boxed{2} \\ \text{TOP} & h_0 \end{array}\right]\end{array}\right]\end{array}\right]$$

c. *everyone*:

$$\left[\begin{array}{ll} \text{PHON} & \langle \textit{everyone} \rangle \\ \text{S L} & \left[\begin{array}{ll} \text{CAT HEAD } \textit{noun} & \\ \text{CONT} & \left[\begin{array}{ll} \textit{ur} & \\ \text{HOLES} & \left\{h_0, h_2\right\} \\ \text{LABELS} & \left\{\boxed{3}\forall y[h_2]\right\} \\ \text{CONSTR} & \left\{\boxed{3} \leq h_0\right\} \\ \text{MAIN} & \boxed{3} \\ \text{TOP} & h_0 \end{array}\right]\end{array}\right]\end{array}\right]$$

The verb *read* in (9a) has two elements on its SUBCAT list. Its CONTENT value is a *ur* which has exactly one hole in its HOLES set. The LABELS set contains the expression $\mathsf{read}'(x,y)$, which specifies the genuine semantic contribution of the verb. There is also one element in the CONSTR set, which is the statement that the expression $\mathsf{read}'(x,y)$ is ordered below the hole h_0 by the relation $\leq$. Finally, h_0 is the TOP value, as it is the highest expression in the UR with respect to the relation $\leq$, and $\mathsf{read}'(x,y)$ is the MAIN value.[5]

In the lexical entry of the quantified NP *something* in (9b), there are two elements in the HOLES set: one hole, h_0, is also the TOP value, the other, h_1, appears as part of the expression in the LABELS set, $\exists y[h_1]$. In this expression, the quantifier $\exists y$ is introduced, but its scope is left underspecified by using the meta-variable h_1. A constraint in the CONSTR set specifies the relative ordering of the expression $\exists y[h_1]$ and the meta-variable h_0 with respect to the relation $\leq$. The lexical entry for *everyone* in (9c) is analogous.

The CONTENT value of phrasal nodes is determined by those of their daughters and the SEMANTICS PRINCIPLE. As is usual in discontinuous semantic systems, the semantic representation of a phrase mainly collects the semantic information of its daughters and adds additional subordination constraints. In (10) we state those parts of the SEMANTICS PRINCIPLE that are needed in order to analyze sentence (1a).

(10) The SEMANTICS PRINCIPLE:

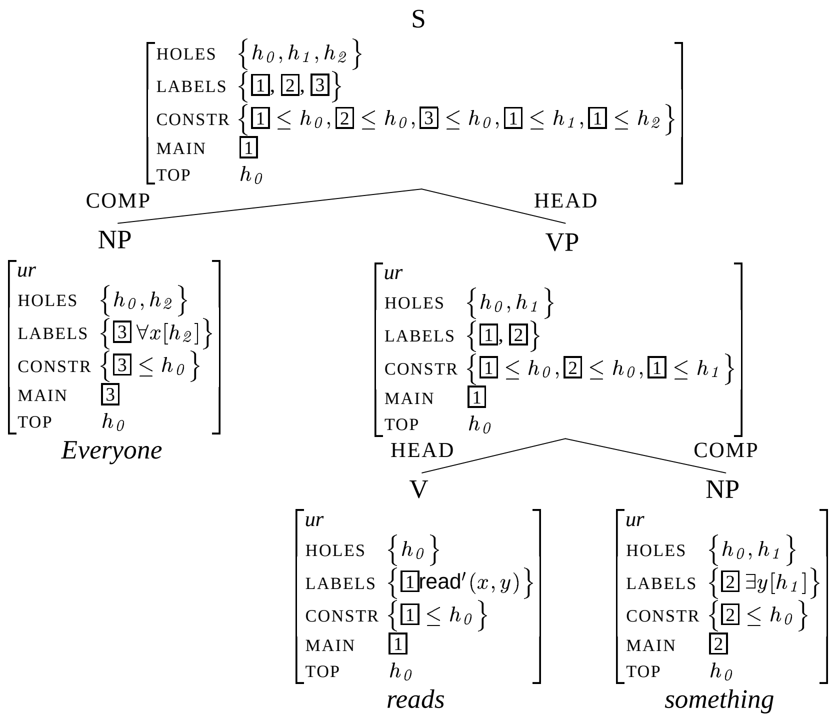
In each phrase:

1. HOLES: is the union of the values at the daughters.
2. LABELS: is the union of the values at the daughters.
3. MAIN: is identical to the head daughter's MAIN value.
4. TOP: is identical at the mother and the daughters.
5. CONSTR: contains exactly
   - all elements of the daughters' CONSTR sets,
   - in *head-complement-phrase*: the head's MAIN value $\leq$ the hole in the complement's MAIN value,
   - ...

The SEMANTICS PRINCIPLE specifies that the HOLES and the LABELS sets of a phrase are the union of the corresponding sets of its daughters. The MAIN value of the phrase is identical to the MAIN value of its head daughter, and the TOP value is assumed to be identical on all signs in a sentence. Thus, in the present architecture, a phrase may not introduce additional holes or labels; all holes and labels are grounded in lexical items. The fifth clause of (10) fixes the phrase's CONSTR value. A phrase contains all the constraints that appear in the daughters' CONSTR values.

[5]As we are only concerned with scope ambiguity, we assume that the identification of the semantic arguments is done lexically. Since the SUBCAT value of the verb contains the *synsem* objects of the arguments, a variable that denotes the respective syntactic arguments is available in the lexical entry and can, thus, directly be used in the semantics of the verb. This mirrors the treatment in Pollard and Sag (1994). See also Halvorsen (1995) for general considerations.

Figure 1: The analysis of the sentence *Everyone reads something.*:



Depending on the kind of phrase, further constraints may be added. In our example sentence, there are only head-complement combinations. For these, the SEMANTICS PRINCIPLE specifies that the MAIN value of the head has to be in the scope of the quantifier that is introduced by the complement. Technically this is expressed by ordering the head's MAIN value below the hole that appears in the CONTENT MAIN value of the complement by the relation $\leq$. As we saw in the lexical entries in (9b,c) this hole corresponds to the scope of the quantifier. In a larger fragment, there is roughly one additional specification in the SEMANTICS PRINCIPLE for each way of syntactic combination.[6]

The effect of the SEMANTICS PRINCIPLE can be seen in Figure 1. It shows the syntactic structure of sentence (1) together with the CONTENT specification at the nodes. The VP has the two holes, h_0 and h_1, in its HOLES set and the two expressions referred to by the tags $\boxed{1}$ and $\boxed{2}$ in its LABELS set. The CONSTR set contains not only the constraints that are part of the UR of the words, but also a third constraint that puts the expression $\mathsf{read}'(x, y)$ in the scope of the existential quantifier ($\boxed{1} \leq h_1$). Analogously, at the S node, the constraint $\boxed{1} \leq h_2$ is introduced, ensuring that the expression $\mathsf{read}'(x, y)$ ends up in the scope of the universal quantifier contributed by the subject.

The resulting CONTENT value at the S node can be interpreted as an underspecified representation in the sense of Bos (1996), namely the one in (4) above. As this UR has exactly the pluggings given in (6), we know that the derived CONTENT value is an underspecified representation of exactly the right readings.

3.1.3 Problematic Data

In Ty2U, the building blocks of a semantic representation are collected together with certain constraints which will, ultimately, restrict the number of scope readings of an utterance. However, complete scope resolution is not part of the grammar, and even for a scopally unambiguous sentence, it might not always be clear from its underspecified repre-

[6] An example is the fragment presented in Richter and Sailer (1999c). Of course, if the constraints to be added depend on the ID schema that licenses the relevant syntactic combination, then these additional clauses of the SEMANTICS PRINCIPLE may better be stated as part of the respective ID schemata. Something similar is done in *constructional HPSG* (Ginzburg and Sag, 2000; Sag, 1997; Riehemann, 1997, 2001). There, phrases may make construction-specific semantic contributions, which means that they may introduce additional holes and labels.

sentation that there exists exactly one reading. This is a natural consequence of having a single semantic representation for scopally ambiguous sentences. On the other hand, leaving possible scopings unresolved makes it impossible to exclude sentences whose ungrammaticality is due to conflicting scoping requirements. The study of those data puts the highly abstract question of whether semantic systems with underspecified denotation such as Ty2U are adequate for theoretical approaches to natural language semantics on an empirical basis. Below we investigate two examples from work on interrogatives and negation more closely.

Example 1: interrogatives in German Beck (1996) discusses the interaction of interrogatives, negation and word order in German. To show the necessity of scope resolution in order to identify ungrammatical constructions in the grammar, we must presuppose some of Beck's theoretical assumptions.

First, an interrogative clause is assumed to be of the semantic type $(st)t$. Its logical form is an expression $\lambda p_{st}.\phi_t$, where ϕ essentially involves the presence of an interrogative operator of the form $p = \lambda w_s.\phi_t$. Second, we adopt a quantificational analysis of interrogative words, following Groenendijk and Stokhof (1982) and many others but in contrast to Ginzburg and Sag (2000). Under Beck's analysis, an interrogative word like *wer* (*who*) is analyzed as an existential quantifier which outscopes an interrogative operator. This is indicated in the simple question in (11).

(11) a. Wer hat Maria geholfen?
who has Maria helped 'Who helped Maria?'

b. $\lambda p_{st}.\exists x[p = \lambda w_s.x$ helped Maria in $w]$

Based on independent observations concerning scope in German, Beck formulates a principle that says that the relative scope of a negation and a quantifier in German are (largely) determined by their relative word order. With these assumptions, Beck predicts the right readings for sentences with multiple interrogative words and negation such as (12).

(12) a. Wann hat wem niemand geholfen?
when has whom nobody helped 'When did nobody help whom?'

b. $\lambda p_{st}.\exists t \exists x[p = \lambda w.\neg\exists y[y$ helped x at $t]]$

As can be seen in the logical form in (12b), the quantifiers contributed by the two interrogative words outscope the interrogative operator ($p = \lambda w.\psi$) and the negation, which is part of the semantic contribution of the n-word *niemand*. Beck also correctly excludes sentence (13).

(13) * Wann hat niemand wem geholfen?
when has nobody whom helped

There, the in situ interrogative pronoun *wem* must take scope over the interrogative operator, just as it does in (12b). The NP *niemand*, which contributes sentential negation, precedes the in situ interrogative and, thus, according to Beck's assumptions, must outscope the quantifier contributed by *wem*. On the other hand, the negation must be in the scope of the interrogative operator $p = \lambda w.\psi$. This leads to conflicting scope requirements, which are made responsible for the ungrammaticality of (13).

In a denotationally underspecified semantic framework such as Ty2U the scope requirements of the semantic operators might be stated, but the fact that there is no successful way to resolve these requirements in the case of (13) would only be detected in an extra-grammatical scope-resolution procedure.

Example 2: n-words in Polish Our second argument is based on data from Polish and English negation, and is quite subtle. We show that whichever of the two options of analyzing the Polish negation particle *nie* one chooses, one is faced with the problem of not being able two capture one set of data within the grammar, if the logical representations are taken from Ty2U.

Hypothesis A says that we have two lexical entries for *nie* (and *not*) which differ in their semantics. Call the two lexical items licensed by them LI$_{1}$ and LI$_{2}$. One of them, assume it is LI$_{2}$, is inappropriate in certain contexts. Under Hypothesis A, however, it turns out to be impossible to exclude the occurrence of LI$_{2}$ in the relevant contexts inside the grammar, since only conflicting scope requirements exclude it and determine that LI_1 is the only possible instantiation of *nie* in those contexts. But scope resolution is extra-grammatical and not part of the grammar.

Hypothesis B says that we have one single lexical entry of *nie* and *not* with an underspecified semantic representation. However, certain Polish words may only occur with the reading of *nie* resolved to one of the two, and our underspecified logical representation no longer permits this distinction in our grammar. Since we have no alternative to Hypothesis A and Hypothesis B, we have to reject Ty2U as our logical representation language if we want to capture the data inside the grammar rather than through extra-grammatical resolution procedures.

We will begin our argument by considering the data and the assumptions that they have elicited. Przepiórkowski and Kupść (1999, p. 216) have shown that the pre-verbal particle *nie* (glossed as NM (Negative Marker)) is systematically ambiguous between eventuality negation and non-eventuality negation (pleonastic or other). Sentences introduced by *omal* (*almost*) such as (14b) constitute contexts for the non-eventuality use of *nie*. As the dual function of *nie* is systematic, it seems at first plausible to leave the particular use of *nie* underspecified in a semantic representation.

(14) a. Jej nie przewróciłem.
her NM I overturned 'I did not knock her over.'

b. Omal jej nie przewróciłem.
almost her NM I overturned 'I almost knocked her over.'

Support for an underspecified treatment of negation comes from data on Positive Polarity Items (PPI). A PPI such as English *already* may occur in non-negated sentences such as (15a). If a sentence contains an eventuality negation, the occurrence of *already* is excluded (Ladusaw, 1980; Linebarger, 1980; Zwarts, 1981).

(15) a. Peter already left.

b. Peter didn't already leave.
("*" under eventuality negation)
(grammatical under meta-linguistic negation)

Example (15b) is only well-formed under a meta-linguistic interpretation of the negation.

The first step of our argumentation is to refute Hypothesis A, which postulates two lexical entries with distinct semantics. Refuting Hypothesis A will naturally steer us towards Hypothesis B which is the one that we have already suggested in response to the data. If we can reject Hypothesis A, Hypothesis B would be strengthened to the point of saying that the negation data can only be treated in Ty2U if we assume the indicated underspecified representation of eventuality and meta-linguistic negation.

First, we adopt Hypothesis A: We assume that eventuality negation and meta-linguistic negation are bound to distinct lexical elements. The purpose of this argument is to show that treating the negation as being expressed by two distinct lexical items will not allow us to exclude the eventuality reading within the grammar. Under Hypothesis A, sentence (15b) receives two potential analyses: One containing eventuality negation, and one containing meta-linguistic negation. In the case of eventuality negation, the sentence can only be grammatical if *already* takes scope outside the negation. This reading, however, is not possible because *already* needs to modify an eventuality that has a result state, whereas *not leaving* lacks such a result state. Therefore, the negated eventuality cannot be in the scope of the adverb. The ungrammaticality of sentence (15b) under eventuality negation is, then, due to a conflict in scoping requirements. As these requirements are not resolved within the grammar, its ungrammaticality is not appropriately accounted for in Ty2U.

Working with Ty2U we must reject the assumption that there exist two distinct lexical elements in the case of eventuality negation and meta-linguistic negation. Therefore, we proceed to Hypothesis B and treat the systematic ambiguity of English *n't* (that we have also observed for the Polish negative marker *nie* in (14)) between eventuality negation and meta-linguistic negation as an instance of underspecification. Then, sentence (15b) receives a single underspecified analysis and the extra-grammatical resolution procedure will correctly provide only the meta-linguistic reading.

Having established the necessity of an underspecified treatment of the negative marker within Ty2U, we can now show that Hypothesis B will still ultimately lead to a conflict. In particular, we will not be able to exclude an ungrammatical sentence by the grammar because of conflicting scope requirements.

Let us return to the Polish examples. In Section 4.2.1 we will show that n-words must co-occur with *nie* in the same clause. The n-word, however, requires an eventuality use of *nie*. This leads to a conflict of requirements in (16).

(16) ?* Omal nikogo nie przewróciłem.
almost nobody NM I overturned

In (16), the n-word requires an eventuality negation interpretation of *nie*, but *omal*, just as in (14b), requires a non-eventuality negation. If, as suggested above, the interpretation of *nie* is not specified in the semantic representation of (16), the conflict that arises from the requirements of *omal* and *nikogo* cannot be detected in the grammar itself. Hypothesis B is therefore refuted as well.

In short, we observe the following situation: If the difference between eventuality negation and meta-linguistic negation is treated via underspecification, the Polish *omal* data cannot be accounted for in the grammar. If, on the other hand, we treat them as distinct, the PPI data cannot be explained within the grammar. As far as we can see, there are no arguments for attributing the oddness of either (15b) or (16) to extra-grammatical factors, while treating the other example inside the grammar. Thus, using Ty2U gets us into a very serious theoretical dilemma.

The conceptual problem caused by the data discussed in this subsection is inherent to approaches of underspecified semantic denotation. For these data, the scoping constraints of the involved elements contradict each other to the extent that even though the sentence is otherwise impeccable, the conflicts lead to ungrammaticality.

3.1.4 Technical Difficulties

As we saw in Section 3.1.3, there are semantic data that are impossible to capture with denotationally underspecified systems inside the logical grammar specification. In order to capture a certain class of data, denotationally underspecified systems must resort to the conceptually unpleasant, if not dubious, move of predicting certain patterns of ungrammaticality by procedures that are intentionally left outside the logic of the grammar. This contradicts established and well-motivated views about the devision of labor in, and the subject matter of the theory of grammar. Besides this serious theoretical shortcoming, denotationally underspecified systems are facing a number of unsolved technical problems.

First, in such systems, scoping constraints are encoded as linguistic entities. Thus, it is necessary to indicate which kinds of constraints exist. In Ty2U we have followed Bos (1996) in allowing only constraints of the kind *"is subexpression of/is in the scope of"*. Egg (1998) assumes 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 *"if l_1 is a subexpression of l_2, then l_3 is a subexpression of l_4"*. To our knowledge, there is no principled discussion of this topic among the proponents of underspecified systems.

The empirical requirements seem to be quite strong, and might even go beyond the kinds of constraints that have been proposed so far. For illustration, consider the ambiguous sentence in (17). In (a)–(d) we indicate available and unavailable (marked $) readings of this sentence.

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

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

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

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

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

The empirical generalization that emerges from these data—which is also incorporated in many analyses in terms of NONLOCAL features (Pollard and Yoo, 1998; Ginzburg and Sag, 2000)—is that if one in-situ interrogative pronoun takes scope over clause C then so does every clausemate in-situ interrogative pronoun. In (a,b), the two in-situ interrogative pronouns are both inside, respectively outside, the scope of the matrix verb. Crucially, in the unavailable reading (17c) *was* (*what*) takes scope over the embedded clause, whereas *wann* (*when*) takes its scope inside the embedded clause. In (17d), *wann* (*when*) has wide scope, and *was* (*what*) takes scope inside the embedded clause. Clearly, such a complex restriction cannot be expressed with the simple kind of subordination constraints that we have assumed for Ty2U so far, nor with any of the kinds of constraints that have been proposed in the literature. Nevertheless this is the kind of restrictions that we would like to state directly as constraints on logical forms at the syntax semantics interface.

Second, well-formedness constraints on underspecified representations are awkward to state. Examples are the requirement that the constraints do not lead to a cyclic plugging (noncyclicity) or that every occurrence of a variable be properly bound (variable binding condition). While the necessity of such constraints has been recognized in the discussion of many underspecified systems such as in Egg (1998, p. 46) and in some versions of the work in progress that presents MRS, the formal language of HPSG adopted in these papers is not expressive enough to support a formalization of the necessary principles within the HPSG grammar. Adopting RSRL there is no principled reason for not stating them formally, but they would have to be quite complex.

Third, even though efforts have been made to characterize the empirical domain that should be handled by underspecification (see for example Pinkal (1999)), there is still no agreement on this issue, as shown by our discussion of eventuality and non-eventuality negation in Section 3.1.3. Two more significant cases come from PP-attachment and ambiguous quantifiers. To illustrate the depth of the problem, let us briefly characterize them as a preliminary conclusion to our discussion.

In classical examples of PP attachment ambiguities such as (18), the same PP can potentially be used to modify entities of different semantic types. So far, there is no proposal that would allow for an underspecification of semantic types. There are, however, semantic analyses in which the PP would modify some individual in both readings (where the VP is assumed to denote an event, which would, ontologically, be treated as an individual). Under such an analysis, an underspecified semantic treatment is in principle available. It would, however, also require a single syntactic analysis for the two readings.[7]

(18) Mary observed a man with a telescope.

It is not clear at all whether syntacticians could agree on assuming a single syntactic structure for both readings.

Another type of example is provided in Frank and Reyle (1995) and Egg (1998). There, the scope of a quantifier is underspecified but not its "reading", i.e., whether it is used distributively or collectively. As a consequence, the question in (19) receives two distinct semantic representations, depending on the reading chosen for the quantifiers.

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

a. Detmar. (collective reading)

b. Detmar helped Manfred; Detmar helped Frank; Frank helped Manfred; . . . . (distributive reading)

For evaluating this proposal independent from practical considerations, it would be necessary to develop criteria for deciding which kinds of underspecification have empirical counterparts and which ones do not.

If there are empirical reasons to give up semantic underspecification for cases of different readings of negation, PP attachment and the collective/distributive distinction of quantifiers, it becomes questionable whether it is really a good tool for taming the "combinatorial explosion" of readings that it was originally proposed for.

3.1.5 Summary

Systems with underspecified denotation raise conceptual questions about the limits of grammar and have technical limitations. While these problems might be considered minor in comparison to the advantages of such systems for computational applications, they cannot be ignored for theoretical considerations.

[7] See Richter and Sailer (1996) and Sailer (1997) for some discussion and a proposal.

3.2 Indirect Representation: Ty2U^P

The main reason for the problems of adequate empirical description of Ty2U is the fact that it is impossible to express in such a system that the subordination constraints are not mutually conflicting. To overcome this, we can move from a denotationally underspecified semantic system to a system which uses the same kind of representations, but specifies the readings as part of the grammar. We refer to this system as Ty2U^P(Ty2U with plugging).

To arrive at such a system, we take all the specifications made for Ty2U, including the lexical entries and the SEMANTICS PRINCIPLE and extend the appropriateness conditions of the sort *ur* in such a way that a new attribute, PLUG(GING), becomes appropriate for *ur*. The value of PLUG on *ur* is a list of *plug* objects. In (20) we state the appropriateness conditions for the sort *plug*.

(20) *plug* HOLE *hole*
LABEL *label*

A *plug* object is interpreted as expressing an identity requirement between a meta-variable (its HOLE value) and a labeled formula (its LABEL) value. In a sentence, we require that the PLUG value expresses a bijection between the elements in the HOLE and the LABEL sets that satisfies the constraints specified in the CONSTR list. As a result, sentence (1) receives one syntactic structure but two semantic representations which differ with respect to their PLUG values. Thus, Figure 1 remains a valid description of the derivation of both readings of sentence (1) in Ty2U^P, but two distinct CONTENT values are in the denotation of the grammar. More precisely, Figure 1 now describes two sentence types in the denotation of the grammar that differ only in their CONTENT values. These CONTENT values discriminate the two readings. In (21) we indicate the PLUG values for the two readings. These encode exactly the pluggings given in (6) above.

(21) a. The ∀∃-reading: $\left[\text{PLUG}\left\langle\begin{bmatrix}\text{HOLE} & h_0\\ \text{LABEL} & \boxed{3}\end{bmatrix},\begin{bmatrix}\text{HOLE} & h_1\\ \text{LABEL} & \boxed{1}\end{bmatrix},\begin{bmatrix}\text{HOLE} & h_2\\ \text{LABEL} & \boxed{2}\end{bmatrix}\right\rangle\right]$

b. The ∃∀-reading: $\left[\text{PLUG}\left\langle\begin{bmatrix}\text{HOLE} & h_0\\ \text{LABEL} & \boxed{2}\end{bmatrix},\begin{bmatrix}\text{HOLE} & h_1\\ \text{LABEL} & \boxed{3}\end{bmatrix},\begin{bmatrix}\text{HOLE} & h_2\\ \text{LABEL} & \boxed{1}\end{bmatrix}\right\rangle\right]$

As the plugging is an explicit part of the semantic representation in Ty2U^P, the descriptive problems discussed in Section 3.1.3 do not arise. On the other hand, all the technical difficulties and questions remain, as they are due to the use of meta-variables and not to the underspecified denotation. Furthermore, the advantage of having a single semantic representation for ambiguous sentences is lost in Ty2U^P. Thus, Ty2U^P, a system with indirect representations, inherits the technical disadvantages of Ty2U but loses its advantages.

3.3 Fully Specified Denotation, Direct and Continuous Representation: LF-Ty2

Lexical Flexible Ty2 (LF-Ty2) is an HPSG-adaptation of *Flexible Montague Grammar* (FMG) as developed in Hendriks (1993). FMG has been very influential as it reduces Cooper Storage mechanisms to derive different scope readings to an independently needed machinery of type shifting operations. Dekker (1993) has shown that a flexible system can also be applied to dynamic phenomena in natural language semantics.

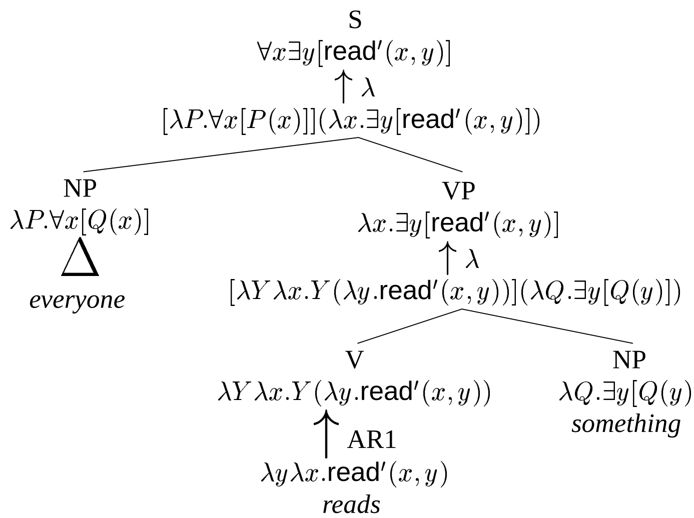
LF-Ty2 differs from FMG in two respects: First, we use Gallin's Ty2 instead of Montague's *Intensional Logic* as semantic representation language. Second, we adopt the proposal of Bouma (1994) to restrict the application of type shifting operations to lexical nodes. The resulting system is fully defined in Sailer (2000) and has been applied in HPSG analyses of Negation in French (Richter and Sailer, 1999a) and Polish (Richter and Sailer, 1999b), and of idiomatic expressions (Sailer, 2000).

In LF-Ty2 the CONTENT value of a sign is taken to be an expression of the semantic representation language, i.e., Ty2. For semantic combination, we assume a basic translation for every word, as given for example in (22).

(22) *read* $\leadsto \lambda y \lambda x.\mathsf{read}'(x,y)$ *everyone* $\leadsto \lambda P.\forall x[P(x)]$ *something* $\leadsto \lambda P.\exists y[P(y)]$

At phrasal nodes, the only way of combining the semantic contributions of the daughters is by *(intensional) functional application*. In order to resolve type conflicts and to account for different scoping possibilities, a set of type

Figure 2: The ∀∃ reading of the sentence *Every student reads some book*:



shifting operations is introduced. For our example, only one of the shifting operations proposed in Hendriks (1993) is needed: *Argument Raising*, illustrated in (23).[8]

(23) Argument Raising:

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

By applying AR, it is possible to change a verb's semantic valence from taking an individual type argument (e) to taking an argument of the type of a quantified NP ($(et)t$). If a transitive verb (of type $e(et)$) combines with a quantified direct object, its type must be shifted by AR to allow for functional application. In Figure 2 the ∀∃-reading of the example sentence is derived by applying AR to the first semantic argument of the verb.

To derive the ∃∀-reading, we need to apply AR to both semantic arguments of the verb. As a result, the verb becomes the semantic functor at all phrasal nodes. The order of the applications of AR, then, specifies the relative scope of the quantifiers: Applying AR first to the x-argument of the verb and then to the y-argument allows us to derive the wide-scope reading for the direct object quantifier. This is shown in Figure 3.

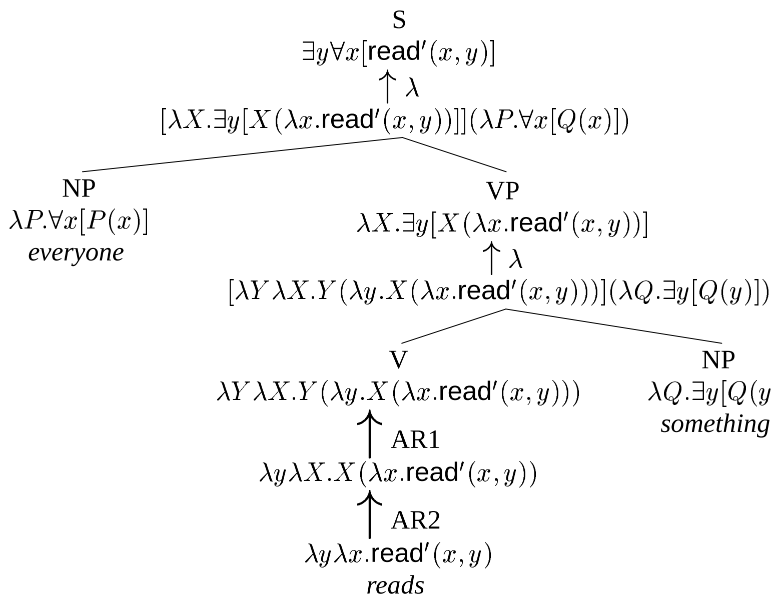
LF-Ty2 does not have the technical or empirical disadvantages of systems such as Ty2U or Ty2U^P. On the other hand, the flexible type shifting machinery leads to a huge number of different derivations for the same readings[9] and requires some experience to work with. In addition, the RSRL encoding of λ-conversion leads to conceptual questions of the ontology of linguistic signs (see the discussion in Sailer (2000, Section 4.2)). LF-Ty2 does not address the concern about the combinatorial explosion of readings at all.

While it is possible to integrate the ideas of FMG seamlessly with the logical approach of constraint-based grammar, the computational questions that first led to the conception of underspecification and to the use of underspecified semantics in HPSG remain. We are thus looking for a theory of semantic representations that is not a mere reconstruction of an existing theory in the form of a logical specification, but a new theory that takes advantage of the genuine possibilities that arise in constraint-based grammar, builds on previous experience with logical grammar specifications, and, if possible, lends itself to efficient computation. Especially for the latter purpose, we hope to learn from the techniques of underspecification that we have studied in connection with Ty2U.

[8]The formulation of AR given here is a simplified version of that in Hendriks (1993). The HPSG formalization in Sailer (2000) follows Hendriks more directly.

[9]Alternatively to the derivation in Figure 2, we could derive the ∀∃-reading by applying AR first to the y-argument of the verb and then to the x-argument.

Figure 3: The $\exists\forall$ reading of the sentence *Everyone reads something*:



4 Discontinuous Representation: LRS

In this section, we will introduce a new semantic meta-theory, *Lexical Resource Semantics* (LRS). The taxonomy of semantic systems in Table 1 helps us to locate LRS with respect to the systems that we have sketched above. Just like LF-Ty2, LRS representations specify individual readings. Doing this, we avoid the problems discussed in Section 3.1.3. For not running into the technical difficulties of indirect systems that we identified in Section 3.1.4, there are no meta-variables in LRS either. Still, LRS exploits the combinatorial simplicity of underspecified systems rather than using λ-conversion and type shifting, which are very cumbersome to integrate into HPSG. Their specification is highly technical and cannot gain anything from the advantages that a logical specification in an expressive description language can offer for formulating linguistic theories elegantly. An important property that LRS shares with Ty2U is the use of discontinuous representations. We will show that this particular combination of properties leads to a straightforward formalization in RSRL and offers some new analytical possibilities.

LRS was first presented in Richter and Sailer (2001a) with an application to interrogatives in German. In Richter and Sailer (2001b) we show how negation in Polish can be analyzed in a natural way using LRS. In the present paper, we first present an LRS analysis of a scopally ambiguous sentence and then move to the discussion of Polish negation in Section 4.2. The following analysis provides a deeper elaboration of how empirical data motivate the architecture of LRS. Our present discussion goes well beyond the motivation of LRS presented in Richter and Sailer (2001b). The fragment of Polish that we will provide contains a new explicit analysis of negation in opaque contexts.

4.1 A Sketch

We introduce a sort *lrs* and assume that the CONTENT value of a *sign* is an object of sort *lrs*. In (24) the appropriateness conditions for this sort are given.

(24) The sort *lrs*

lrs	EX(TERNAL-)CONT(ENT)	*me*
	IN(TERNAL-)CONT(ENT)	*me*
	PARTS	*list(me)*

If you compare the sort *lrs* to the sort *ur* in (7), the attributes EXCONT and INCONT in (24) correspond directly to the attributes TOP and MAIN of the sort *ur* and fulfill basically the same function. In particular, the INCONT value is seen as expressing the genuine semantic contribution of a word, and the EXCONT value is the expression that indicates the overall logical form of a phrase.[10] In the attribute PARTS the meaning contributions of the words in a sentence are collected. Thus, it corresponds to the attribute LABELS on *ur* objects. We will introduce a principle that guarantees that in an utterance, the EXCONT value is exactly composed of the expressions in the PARTS list. As we do not use meta-variables, there is no need to represent holes or constraints in LRS. Thus, LRS is also ontologically simpler than Ty2U. To illustrate how LRS works, consider the scopally ambiguous sentence in (25), which is slightly more complex than our previous example from (1).

(25) Every student reads some book
$\forall x[\mathsf{student}'(x) \rightarrow \exists y[\mathsf{book}'(y) \wedge \mathsf{read}'(x,y)]]$
$\exists y[\mathsf{book}'(y) \wedge \forall x[\mathsf{student}'(x) \rightarrow \mathsf{read}'(x,y)]]$

To derive the two readings of the sentence, we shall first present the lexical specification of the words that occur in the sentence and, then, indicate the combinatorial principles.

Lexical specifications: The following semantic specifications are assumed in the lexical entries of the words:

(26) Lexical specification of the CONTENT values:

a. *read*: $\begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{1}\,(\mathsf{read}'y)x \\ \text{PARTS} & \left\langle x, y, \boxed{1}, \boxed{1a}\mathsf{read}'y, \boxed{1b}\mathsf{read}' \right\rangle \end{bmatrix}$

b. *book*: $\begin{bmatrix} \text{EXCONT} & \begin{bmatrix} \textit{quantifier} \\ \text{VAR}\ y \end{bmatrix} \\ \text{INCONT} & \boxed{2}\mathsf{book}'(y) \\ \text{PARTS} & \left\langle y, \boxed{2}, \boxed{2a}\mathsf{book}' \right\rangle \end{bmatrix}$ *student*: $\begin{bmatrix} \text{EXCONT} & \begin{bmatrix} \textit{quantifier} \\ \text{VAR}\ x \end{bmatrix} \\ \text{INCONT} & \boxed{3}\mathsf{student}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{student}' \right\rangle \end{bmatrix}$

c. *some*: $\begin{bmatrix} \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{4}\exists y[\alpha \wedge \beta] \\ \text{PARTS} & \left\langle y, \boxed{4}, \boxed{4a}\,[\alpha \wedge \beta] \right\rangle \end{bmatrix}$ *every*: $\begin{bmatrix} \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{5}\forall x[\gamma \rightarrow \delta] \\ \text{PARTS} & \left\langle x, \boxed{5}, \boxed{5a}\,[\gamma \rightarrow \delta] \right\rangle \end{bmatrix}$

In the lexical entry of *read* in (26a), the EXCONT value is left unspecified. The only requirement is that it be an expression of Ty2. The INCONT value is the semantic contribution of the verb together with the argument linking, just as assumed for Ty2U in (9a). In (26a), however, we have written $(\mathsf{read}'y)x$ instead of $\mathsf{read}'(x,y)$ to make its term structure explicit. This expression has as its subexpressions the expressions x, y, read', $\mathsf{read}'y$ and $(\mathsf{read}'y)x$. It should be noted that the PARTS list contains exactly all of its subexpressions.

The INCONT and the PARTS specifications in the lexical entries of the nouns *book* and *student* in (26b) are analogous to those for *read*. For the EXCONT value we do, however, specify that there will be some quantifier that binds the variable which is also used in the INCONT value.

In the lexical entries of the quantifiers in (26c) the INCONT value is not a description which uniquely identifies an expression of Ty2 (up to variable names), but a description which only determines parts of the expression that is being described. In the case of *every*, it is specified that the INCONT value is an expression that starts with a universal quantifier over a variable x that has immediate scope over an implication. Nothing is said about the further subexpressions. In (26c) we use lower case Greek letters for subexpressions that are not further described in the given lexical entries. These letters are formally nothing but tags, but for notational clarity we avoid tags when we use the term notation for expressions of Ty2 instead of the AVM notation. In (27) a pure AVM notation of the CONTENT specification of *every* is given.

[10]In former versions of LRS (Richter and Sailer, 2001a,b; Bouma, 2003), the attribute names TOP and MAIN are used. The names EXCONT and INCONT are borrowed from Kasper (1997). This change in terminology is motivated by the strong similarity of the use of these attributes in LRS and in Kasper's approach to recursive modification.

(27) $\begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{5} \begin{bmatrix} \textit{universal} & \\ \text{VAR} & x \\ \text{SCOPE} & \boxed{5a}\ \textit{implication} \end{bmatrix} \\ \text{PARTS} & \left\langle \boxed{5}, \boxed{5a}, x \right\rangle \end{bmatrix}$

If we look at an expression as a tree where the node labels specify the kind of syntactic combination of the subexpressions, we can say that the PARTS list of *every* contains exactly those nodes of the tree of the INCONT value whose labels are specified, i.e., the variable, the quantifier and the implication.

Combinatorial System We first impose two well-formedness conditions on *lrs*: the INCONT PRINCIPLE (IContP) in (28) and the EXCONT PRINCIPLE (EContP) in (29). Then, in (30) we specify the relevant parts of the SEMANTICS PRINCIPLE.

(28) The INCONT PRINCIPLE (IContP):
In each *lrs*, the INCONT value is an element of the PARTS list and a component of the EXCONT value.

The IContP ensures that the INCONT value of a word appear as an element of the PARTS list. Consequently, the basic semantics of a word is introduced as part of the semantic contribution of the word to the overall logical form. Furthermore, the INCONT value is required to be a subexpression of the EXCONT value. This binds the syntactic and semantic properties of a sign closer together: The EXCONT value is identical throughout a head projection and it is interpreted as the logical form of a sign. Thus, the IContP guarantees that the basic semantics of the lexical syntactic head is part of the logical form of the projection of this head.

(29) The EXCONT PRINCIPLE (EContP):

1. In every phrase, the EXCONT value of the non-head daughter is an element of its PARTS list.
2. In every utterance, every subexpression of the EXCONT value of the utterance is an element of its PARTS list, and every element of the utterance's PARTS list is a subexpression of the EXCONT value.

The EContP sets further constraints on the EXCONT value of a sign. It consists of two clauses, one for phrases and one for utterances. The first clause specifies that the EXCONT value of a completed head projection must be one of the elements of this sign's PARTS list. This is encoded in the EContP in an indirect way: We can only see that a head projection is completed if the sign that we consider is an utterance (clause 2), or if it combines with an other sign to a phrase, where the resulting phrase is no longer a head projection of the sign under consideration (clause 1). In combination with the IContP, clause 1 of the EContP gives us a good general specification of the EXCONT value of a completed head projection: It must be an expression that was part of the semantic contribution (the PARTS list) of some dominated sign, and the expression must contain the basic semantics of the lexical head as a subexpression.

The second clause of the EContP states that in an utterance, every subexpression of the logical form, i.e., the EXCONT value, has been contributed by some lexical element that is dominated by the utterance; and every bit of every expression on the PARTS list occurs somewhere in the semantic representation of the utterance. This ensures that no contribution by any lexical element in the utterance gets lost.

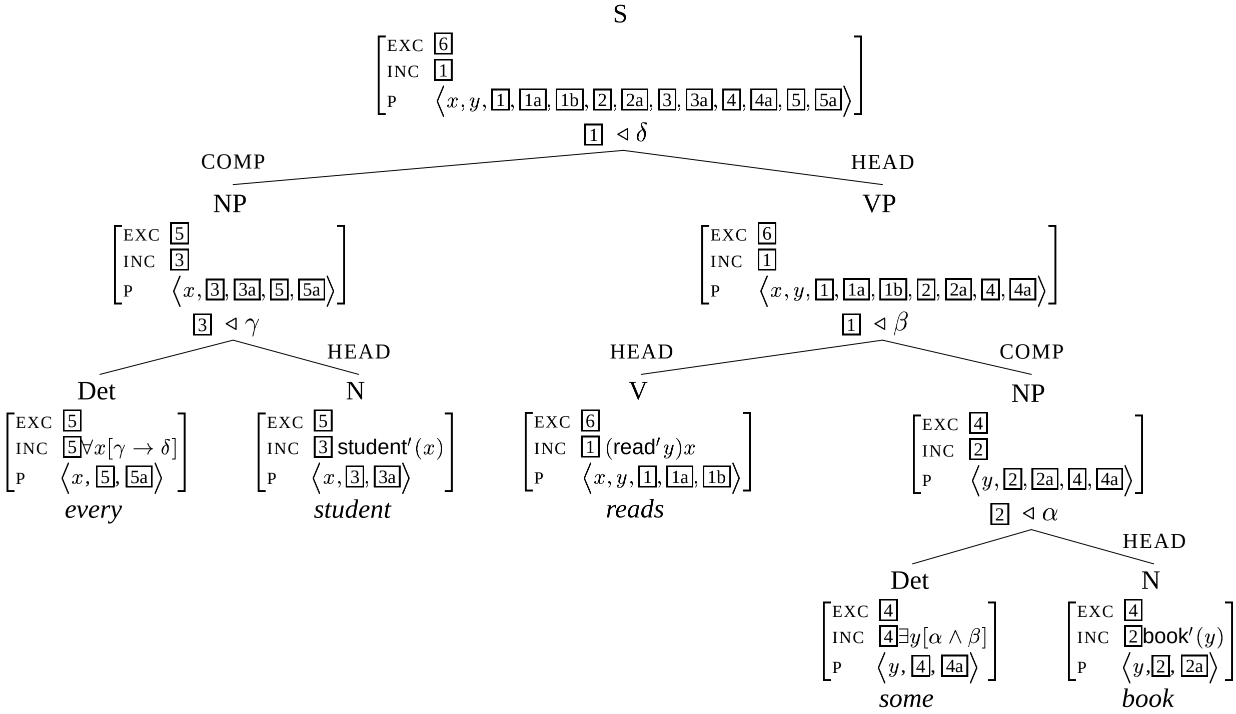
The next important principle is the SEMANTICS PRINCIPLE (SP), given in (30).

(30) SEMANTICS PRINCIPLE
In each *headed-phrase*,

1. the EXCONT value of the head and the mother are identical,
2. the INCONT value of the head and the mother are identical,[11]
3. the PARTS value contains all and only the elements of the PARTS values of the daughters,
4. the following conditions hold:

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

Figure 4: The LRS analysis of the sentence *Every student read some book.*:



(a) if the nonhead is a quantifier then its EXCONT value is of the form $Qx[\rho \circ \nu]$[12], the INCONT value of the head is a component of ρ, and the EXCONT values of both daughters are identical,

(b) if the non-head is a quantified NP with an EXCONT value of the form $Qx[\rho \circ \nu]$, then the INCONT value of the head is a component of ν,

(c) ...

The SP for LRS is very similar to that for Ty2U in (10). Just as was assumed there, the INCONT and the EXCONT values are shared along the head projection line. The PARTS list of a phrase, analogously to the LABELS list in Ty2U, is simply a concatenation of the PARTS lists of the daughters. Clause 4 of the SP is parallel to the condition on the CONSTR list of Ty2U. We do, however, give a more refined syntactic and semantic analysis of quantified NPs, where clause (4a) treats the combination of a quantifier with a head noun and (4b) the combination of a quantified NP argument with its syntactic head. More importantly, we do not introduce constraints on some CONSTR list but instead we can impose the scoping constraints directly on the expressions in the denotation of the grammar.

In Figure 4 the syntactic and semantic derivation of the ambiguous sentence in (25) is given. At the nodes, we give the CONTENT values in AVM notation and, where appropriate, we specify the constraints that are imposed by the SP. For this purpose, we use the symbol " $\triangleleft$ " for the relation "*is a component/subexpression of*". The tags used in this figure are those of the lexical entries in (26). At the two NP nodes, clause (4a) requires the INCONT value of the head ($\mathsf{student}'(x)$ and $\mathsf{book}'(y)$) to be in the restrictor of the universal respectively the existential quantifier ($\boxed{3} \triangleleft \gamma$ and $\boxed{2} \triangleleft \alpha$). At the VP and the S nodes, it is clause (4b) that brings the INCONT value of the verbal head ($\mathsf{read}'(x, y)$) in the scope of both quantifiers: $\boxed{1} \triangleleft \beta$ at the VP node, and $\boxed{1} \triangleleft \delta$ at the S node.

As an effect of the EContP, we can give the EXCONT value of the non-heads in all phrases: At the NP nodes, the EXCONT value of the determiners is identical to their INCONT values, as these are the only elements in the PARTS lists

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

that contain the INCONT value as their subexpression. At the VP node (and, analogously, at the S node), the EXCONT value of the NP *some book* must be identical to that of the determiner: this is the only element in the NP's PARTS list that satisfies the condition expressed in the lexical entry of *book*, i.e., that the EXCONT value be a quantifier that binds the variable y.

At the S node, the second clause of the EContP applies, i.e., the expressions in the PARTS list specify exactly the nodes in the term tree for the resulting logical form. There are, just as was the case for Ty2U, exactly two possible term trees that are compatible with the descriptions that are given for the nodes in the term tree in the lexical entries and that also fulfill the subexpression requirements (i.e., the requirements on domination in the term tree) given by the SP: either $\boxed{6}$ is identical to $\boxed{5}$ to obtain the $\forall\exists$-reading, or $\boxed{6}$ is identical to $\boxed{4}$ for the $\exists\forall$-reading.

In our exposition of LRS, we frequently pointed to the parallels with Ty2U. Regarding crucial properties of its architecture, LRS is technically very similar to Ty2U. The major difference between the two systems lies in the levels at which underspecification techniques are employed. In Ty2U, the denotation of the CONTENT value of an utterance is a UR which receives a special interpretation. The most important means of underspecification are meta-variables in the logical representation language, which leave holes to plug in other pieces of logical expressions. The possible pluggings depend on a partial order on meta-variables and constraint labels. In the case of LRS, underspecification is lifted to the specification language, RSRL, of the grammar formalism. The expression denoted by the EXCONT value of the grammar are fully specified expressions of the logical representation language. Constraints on these logical expressions are collected in the syntactic tree. Any restriction that can be expressed in RSRL may serve as a constraint on the logical forms, but the most important kind of constraint are componenthood constraints, symbolized as $\triangleleft$, and technically defined as an RSRL relation.

4.2 Case Study: Negative Concord in Polish

In this section, we conduct a case study of Negative Concord (NC) in Polish. We continue the discussion begun in Richter and Sailer (2001b) and enrich its empirical coverage and analytical devices. We chose the NC phenomenon, because negation is a scope-bearing element which interacts with quantification. This property makes a treatment in terms of underspecification very attractive. In addition, much work within HPSG has been devoted to the empirical and to the theoretical side of Polish negation (Przepiórkowski and Kupść, 1997a,b, 1999; Kupść and Przepiórkowski, 1998; Przepiórkowski, 1999b; Richter and Sailer, 1999b). It may thus be considered a well-studied phenomenon for which we find a stock of established theories. Further major studies on the topic are Witkoś (1996) and Błaszczak (1998, 1999). The insights gained in these papers form the empirical background of our discussion. Our case study of NC in Polish will illustrate the specific techniques available in LRS for analyzing empirical phenomena in natural language semantics.

The basic data of Polish NC are quickly summarized. Sentential negation is expressed by the pre-verbal particle *nie*, glossed as NM. *Nie* has been analyzed as a verbal prefix (see Kupść and Przepiórkowski (1998)).

(31) Janek nie pomaga ojcu.
Janek.NOM NM helps father.DAT 'Janek doesn't help his father.'

If an n-constituent appears in a sentence, the presence of *nie* is obligatory to achieve sentential negation in Polish (Przepiórkowski and Kupść, 1999).[13]

(32) a. Janek *(nie) pomaga nikomu.
J..NOM (NM) helps nobody.DAT 'Janek doesn't help anybody.'

b. Nikt *(nie) pomaga ojcu.
nobody.NOM (NM) helps father.DAT 'Nobody helps his father.'

c. Janek nigdy *(nie) pomaga ojcu.
Janek.NOM never (NM) helps father.DAT 'Janek never helps his father.'

[13] See Przepiórkowski and Kupść (1999) and Błaszczak (1998) for instances of a non-negative use of n-words, as in the English 'She is a nobody.' They will be ignored throughout this paper.

More than one n-constituent may appear in a sentence (33). Pre-verbal *nie* is obligatory for sentential negation, and, even if there are several n-words, no double negation effect arises.

(33) a. Nic nikomu *(nie) powiedziałem.
nothing.GEN nobody.DAT (NM) I-told 'I didn't tell anybody anything.'

b. Nikt nigdy nikogo niczym *(nie) uszczęśliwił.
'Nobody has ever made anybody happy with anything.'

Double negation does occur, however, if more than one verb is preceded by an NM (Witkoś, 1996, p. 82). In (34), double negation is the only available reading.

(34) Tomek nie może nie znać Marii.
Tomek.NOM NM may NM know Maria.GEN
'It is not the case that it is possible that Tomek does not know Maria.'

When it comes to the discussion of NC in linguistic literature, there are two major questions that need to be answered for n-words: (i) Are they inherently quantificational? and (ii) Are they inherently negative? Depending on the language, the answers to these questions may be different.[14] In the present paper, we will be primarily concerned with the second question. In the work of Joanna Błaszczak and in Richter and Sailer (1999b), it is concluded that n-words are not negative. In the present paper, we want to question this conclusion and argue that at least in the case of Richter and Sailer (1999b), the conclusion was caused by the semantic system assumed, namely LF-Ty2. It is not consistent with a natural interpretation of the data. We can, then, show that using LRS we get an empirically more satisfactory and technically more elegant analysis.

4.2.1 Data

The goal of this section is to show that Polish n-words are genuinely negative. We collect the positive evidence for this negativity and refute counterarguments found in our body of literature.

The most direct evidence for the inherent negative character of Polish n-words comes from contexts in which these n-words are the only possible overt marker of negation. In Przepiórkowski and Kupść (1999) and Błaszczak (1999) three contexts are mentioned: short answers (35), *either or*-construction (39) and *jak*-comparatives (40).

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

In these cases, the answer is interpreted as negative. Had there been a non-negative answer, such as *Piotra* (*Peter.*ACC), a negative interpretation of the answer would be impossible. It had been assumed, for example in Błaszczak (1999), that short answers involve ellipsis. There are, however, problems with the elliptic explanation, at least for some data. In negated sentences, a direct object appears in genitive rather than accusative case in Polish. This phenomenon is called *Genitive of Negation* (GoN) and has been investigated in detail in Przepiórkowski (1999a). In (35) the n-word *nikogo* can be both accusative or genitive, but the feminine form of the n-word *żaden* allows us to distinguish these cases. As shown in (36a), in a short answer both the accusative and the genitive form of the n-word are possible. In the long answer, however, the genitive is obligatory.

(36) Ile przeczytałeś książek?
How many you read books 'How many books have you read?'

a. Żadnej./ Żadną.
None.GEN./ None.ACC.

b. Nie przeczytałem [żadnej kziążki]/ *[żadną książkę].
NM I read [no book].GEN/ [no book].ACC.

[14] See Giannakidou (1998) for an overview.

It might be possible to work out an elliptic analysis for the cases with GoN. Such an analysis is, however, implausible for the accusative answers. This is strengthened by the following observation. The minimizer *słowo* (*word*) in Polish is a negative polarity item. In a full answer, the negative marker *nie* occurs and the minimizer appears in genitive case, i.e., there is an instance of GoN. In a short answer, however, the NPI must occur in genitive as well. An accusative form is ungrammatical if a minimizer interpretation is intended.

(37) Powiedział coś? *Słowo./ Słowa./ Słowa nie powiedział.
he said something? Word.ACC/ Word.GEN/ Word.GEN NM he said
'Did he say something/anything? Not even a word./ He did not say even a word.'

This is a clear contrast between minimizers and n-words. If we postulate an elliptic analysis for the cases with GoN, we correctly predict that both n-words and minimizers are grammatical in short answers. The impossibility of an accusative minimizer in short answers shows that the assumption of an elided negation in this context must be rejected. The fact that accusative n-words may occur, thus, indicates that n-words must be inherently negative.

Błaszczak (1999, p. 145) points to an other set of data to challenge the assumption of inherent negativity in short answers. In existential sentences in Polish, the subject appears in nominative case if the sentence is not negated and in genitive if the sentence is negated. According to her judgments, the question in (38) cannot be answered with a simple n-word, be it in nominative or genitive case. Her argument is that the underlying form that would be the basis for the elision is too different from the possible short answers.

(38) a. Był tam ktoś?
Was there someone? 'Was there anybody?'

b. Błaszczak (1999): *?Nikt./ *?Nikogo./ Nie, nie było tam nikogo.
Nobody.NOM/ Nobody.GEN/ No NM was there nobody.GEN
'Was there anybody? Nobody/ No, there was nobody.'

c. our informants: ?Nikt./ *Nikogo./ Nie, nie było tam nikogo.
Nobody.NOM/ Nobody.GEN/ No NM was there nobody.GEN

There are, however, two problems with this argument. First, the question in (38) is a yes-no-question, so an n-word is not the most natural way to give a short answer. We think that this observation already accounts for Błaszczak's judgments. Second, we found informants who accepted n-words as short answers to the yes-no question. For these informants, however, only the nominative case was acceptable in the short answer. These judgments are given in (38c). Clearly the case in the short answer is different from that in the full answer which, for all informants, must be genitive. This, again, suggests that the negative interpretation of the answer stems from the n-word rather than from an underlying and elided base.

We conclude that at least for examples in which the short answer is not in genitive, it would be highly stipulative and even problematic, if we assumed an underlying (elided) negation. Instead the n-word is the only possible source of a negation.

The picture we just found for short answers reappears in *albo* (*either-or*) constructions as in (39). In (39a) the part of the sentence after the second *albo* does not contain a verb. If an n-word occurs there, this part is interpreted negatively. Following *albo*, there can be an accusative or a genitive form of the n-constituent. Just as we saw with answers, *albo* can also be used to conjoin entire clauses. In the latter case the n-word must be in GoN (39b).

(39) a. Chcę poślubić albo ją, albo nikogo/ [żadnej innej]/ [żadną inną].
I want marry either her or nobody/ [no other].GEN/ [no other].ACC

b. Chcę poślubić albo ją, albo [żadnej innej]/ *[żadną inną] nie chcę poślubić.
I want marry either her or [no other].GEN/ [no other].ACC NM I want marry

Similar data have been given for *jak* comparatives as in (40) in Przepiórkowski and Kupść (1999) or Błaszczak (1999).

(40) a. Kocham ją jak [żadną inną]/ ?[żadnej innej].
I love her.ACC as [no other].ACC/ [no other].GEN
'I love her more than (I love) any other (girl).'

b. Kocham ją jak *[żadną inną]/ [żadnej innej]/ nie kocham.
I love her.ACC as [no other].ACC/ [no other].GEN NM I love

Finally, in Polish there are some contexts that license an expletive occurrence of *nie*. Crucially, in those contexts, n-words are not allowed. We already saw an example of this in (16) in Section 3.1.3 above. Przepiórkowski and Kupść (1999) and Błaszczak (1999) discuss further contexts such as clauses introduced by *dopóki* (*until*) or by *zanim* (*before*). It should be noticed that there is a lot of inter-speaker variation with respect to the possibility of expletive *nie*. The data and the judgments are given according to Błaszczak (1999).

(41) a. Zostaniesz w domu, dopóki nie napiszesz wypracowania/ *niczego/ czegoś.
you will stay at home until NM you will write essay.GEN/ nothing.GEN/ something.GEN
'You will stay at home until you write (have written) the essay/ something.'

b. Zrobię to, zanim Maria/ *nikt/ ktoś (nie) przyjdzie.
I will do this before Maria/ nobody/ somebody NM will come
'I will do it before Maria/ somebody comes.'

In these contexts, the presence of *nie* triggers GoN, but as noticed in Przepiórkowski and Kupść (1999), n-words are not possible. If we assume that n-words contribute an eventuality negation, the data follow right away as there is no such eventuality negation involved in the sentences.

Non-negative n-words? There are a few remaining arguments against the inherent negativity of n-words that we would like to address next. All of these arguments are indirect and concern the interaction of different elements in a clause rather than the n-words themselves.

The most prominent argument is that if a clause is negated, the pre-verbal particle *nie* is obligatory. Thus, this element should be analyzed as inherently negative. Then, as we saw in the data in (31)–(33), independent of the number of n-words that occur in the clause additionally, the logical form only contains one negation. There is, thus, a minimality argument that the n-words need not contribute negation. Furthermore, in most systems of combining the logical forms of syntactic daughters at their mother phrases, there is no natural mode of combination that would "get rid" of the superfluous negations that are introduced by multiple n-words with inherent negativity. These considerations were the major reasons for abandoning a negative analysis of n-words in Richter and Sailer (1999b).

A more subtle version of this argument has to do with what can intervene between the negation and the scope of the n-word. As can be seen in the examples in (42), the quantifier that is related to the n-word is not in the immediate scope of the negation "$\neg$". Instead, an opaque operator is intervening that is introduced by an opaque verb in (a) or by a modal in (b).[15]

(42) a. Janek nie szuka żadnego jednorożca.
Janek NM seeks no unicorn
$\neg \mathsf{seek}'_{@}(j, \lambda @ \lambda P.\exists x[\mathsf{unicorn}'_{@}(x) \wedge P_{@}(x)])$

b. Janek nie może nic czytać.
Janek NM can nothing read 'Janek is not allowed to read anything.'
$\neg \mathsf{may}'_{@}(j, \lambda @.\exists x[\mathsf{read}'_{@}(j, x)])$

In frameworks like Flexible Montague Grammar, which we have captured in HPSG as LF-Ty2, there is no obvious way to account for these data if the negation is analyzed as part of the semantic contribution of the n-word. In LRS, where lexical items may specify discontinuous logical contributions to the meaning of expressions, this is not

[15] We write @ for the distinguished world variable of Groenendijk and Stokhof (1982). We say more about our notation below (page 28) when we discuss intensional and modal verbs. For the present argument, these notational details are irrelevant.

a problem and cannot be constructed as an argument for the inherent negativity of n-words. Their scope-bearing semantic contribution is simply realized non-discontinuously in the overall expression to which they belong.

A strong argument against the negativity of n-words could be constructed by adducing a non-negative environment that would still license n-words. On the search for relevant examples, it has been noticed in Przepiórkowski and Kupść (1999) and elsewhere that there is one environment that allows for the occurrence of n-words without pre-verbal *nie*, namely the complement of *bez* (*without*), as shown in (43).

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

It has, however, been argued that *bez* really expresses a negation in its logical form. If this is true, sentence (43) is of course no longer an instance of a non-negative licensing context for n-words.

In other, clearly non-negative contexts n-words are excluded in Polish. For illustration, we quote examples of yes-no questions (44) and *if* clauses (45) from Błaszczak (1999, p. 128f.). These contexts are known to be non-negative but still able to license some negative polarity items, as indicated by the occurrence of *any* in the English translation.

(44) a. * Czy znalazłeś tam nikogo?
whether you find there nobody.ACC 'Did you find anyone there?'

b. Nie znalazłeś tam nikogo, prawda?
NM you find there nobody.GEN true 'You haven't found anyone there, right?'

(45) a. * Jeżeli nikt przyjdzie, daj mi znać.
if nobody comes let me know 'If anyone comes, let me know.'

b. Jeżeli nikt nie przyjdzie, daj mi znać.
if nobody NM comes let me know 'If nobody comes, let me know.'

Before ending our overview of the relevant facts of Polish, we have to mention a data pattern that has not been noticed yet, but is problematic for our account. In the context of the data concerning expletive *nie* in (41), Błaszczak (1999) gives the following example:

(46) Boję się, żeby on/ *nikt/ ktoś nie przyszedł.
I fear REFL that he/ nobody/ somebody NM came
'I am afraid that he/ somebody will/might come.'

According to Błaszczak, an occurrence of an n-word in this sentence is not possible. For some speakers, expletive *nie* seems to be excluded in sentence (46) in the first place. Among those speakers that accept expletive *nie*, however, there is a group that also allows for an expletive interpretation of *nikt*.[16] This group of speakers is not mentioned in Błaszczak (1999). However, they provide the only instances of a non-negative use of n-words that we came across so far. While further research is needed on this phenomenon, it is clear that the occurrence of expletive *nie* is lexically governed and subject to a lot of variation among speakers. The extraordinary behavior of the verbs *bać się* and *obawiać się* (both meaning *fear*) also calls for a lexical treatment.

Conclusion Our discussion of the arguments for a non-negative character of Polish n-words demonstrated that they draw their force from specific assumptions on the way the logical forms of a phrase are derived from the logical forms of its daughters, and what the logical forms look like. The massive evidence for the inherent negative character of n-words calls exactly these underlying assumptions about appropriate techniques for combining the logical contributions of words and phrases into question. We will demonstrate that in LRS the technical problems that preclude a negative analysis of n-words in frameworks such as LF-Ty2 do not arise.

[16]We are grateful to Anna Mlynarczyk for pointing out this judgment pattern to us.

Figure 5: The analysis of *Nikt nie przyszedł* (*Nobody came*):

$$\begin{bmatrix} \text{EXCONT} & \boxed{0} \\ \text{INCONT} & \boxed{1} \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}, \boxed{2}, \boxed{3}, \boxed{3a}, \boxed{4}, \boxed{5}, \boxed{5a} \right\rangle \end{bmatrix} \qquad \begin{array}{l} \& \; \boxed{1} \lhd \alpha \\ \& \; \boxed{2} \lhd \boxed{0} \\ \& \; \boxed{5} \lhd \beta \\ \& \; \boxed{3} \lhd \gamma \\ \& \; \boxed{1} \lhd \delta \end{array}$$

$$\begin{array}{cc} \textit{Nikt} & \textit{nie przyszedł} \\ \begin{bmatrix} \text{EXCONT} & \boxed{5}\, \exists x[\gamma \wedge \delta] \\ \text{INCONT} & \boxed{3}\, \mathsf{human}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{human}', \boxed{4}\neg\beta, \boxed{5}, \boxed{5a}[\gamma \wedge \delta] \right\rangle \end{bmatrix} & \begin{bmatrix} \text{EXCONT} & \boxed{0} \\ \text{INCONT} & \boxed{1}\, \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}', \boxed{2}\neg\alpha \right\rangle \end{bmatrix} \end{array}$$

4.2.2 Analysis

In this section, we will provide an analysis of Polish n-words as contributing an existential quantifier and a negation to the overall logical form of an utterance.

Let us first consider the semantic contribution of *nie*-prefixed verbs. We do not go into the internal structure of such combinations, but implicitly assume an analysis along the lines of Kupść (2000).

(47) Description of the word *nie przyszedł* (*NM came*):[17]

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nie przyszedł} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \boxed{0} \\ \text{INCONT} & \boxed{1}\, \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}', \boxed{2}\neg\alpha, \right\rangle \end{bmatrix} \end{bmatrix} \qquad \begin{array}{l} \& \; \boxed{1} \lhd \alpha \\ \& \; \boxed{2} \lhd \boxed{0} \end{array}$$

The semantic contribution of the verb *nie przyszedł* consists of all the components of the expression $\mathsf{come}'(x)$, just as would be the case for the non-negated form *przyszedł*. In addition, there is a negation ($\neg\alpha$) included in the PARTS list. This can be considered the semantic contribution of the prefix *nie*. The first constraint at the right of the AVM states that the semantic contribution of the verb is in the scope of this negation. According to the second constraint, the negation must be a subexpression of the EXCONT value of the verb. Due to the EContP, this constraint guarantees a clause bound interpretation of the negation.

In (48) we have sketched the lexical entry of the n-word *nikt* (*nobody*). As can be seen, we treat this element just like the existentially quantified NP *some book* above: the INCONT value contains the restriction to humans and the EXCONT value contains the quantifier. In addition, the PARTS list also contains a negation ($\neg\beta$) which is required to have scope over the EXCONT value of the n-word. Note that, in contrast to the negation on the PARTS list of *nie przyszedł*, this negation may scope outside of the head projection of the lexical entry. The crucial difference between the lexical entries lies in the different componenthood constraints between the EXCONT value and the negation on the PARTS list.

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

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nikt} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \boxed{5}\, \exists x[\gamma \wedge \delta] \\ \text{INCONT} & \boxed{3}\, \mathsf{human}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{human}', \boxed{4}\neg\beta, \boxed{5}, \boxed{5a}[\gamma \wedge \delta] \right\rangle \end{bmatrix} \end{bmatrix} \qquad \begin{array}{l} \& \; \boxed{5} \lhd \beta \\ \& \; \boxed{3} \lhd \gamma \end{array}$$

[17]A word on our notation of the languages of RSRL might be necessary here: In order to distinguish between the logical description language and the expressions of the object language of logical representations, Ty2, we use the symbols $\&$ for conjunction, $\Rightarrow$ for implication, and **A** for the universal quantifier of the constraint language of RSRL. That is not the standard notation for these logical symbols in RSRL.

Figure 5 shows the structure of a simple negated sentence that consists just of the words *nikt* and *nie przyszedł*. To the right, we list the subordination constraints that stem from the description of the words and the SEMANTICS PRINCIPLE.

Given this structure, our principles license three possible values for the EXCONT attribute of the sentence. They are listed in (49). Two of the readings depend on whether the negation contributed by the subject has scope over that contributed by the negated verb (a) or the other way around (b). Finally, (49c) expresses the situation where the two negations are identical, i.e., where the expressions referred to by the tags $\boxed{2}$ and $\boxed{4}$ are identical. In fact, this is the only possible reading of the example sentence.

(49) a. $\neg\neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{4} \lhd \boxed{2} = \boxed{0})$ $= \exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$
b. $\neg\exists x\neg[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{2} \lhd \boxed{4} = \boxed{0})$ $= \forall x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$
c. $\neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)]$ $(\boxed{2} = \boxed{4} = \boxed{0})$

To exclude the empirically impossible (a) and the (b) readings, we posit the following constraint for Polish:

(50) The NEGATION COMPLEXITY CONSTRAINT (NCC):

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

The NCC is language-specific. Since Polish is an obligatory NC language, there may be at most one sentential negation. This constraint is an adaptation of complexity constraints on semantic representations as they were for example postulated for French in Corblin (1994, 1995) and for other Romance languages in Corblin and Tovena (2001). In these papers it has been observed that there is not only a general strategy within the languages of the world to minimize the number of negations that appear in the semantic representation of a clause, but that this strategy might even get grammaticalized at a certain threshold. While for French this threshold is postulated to be two negations, for Polish only one negation is admissible.

The NCC achieves the same effect as the *negation absorption* operation of Haegeman and Zanuttini (1991), but, whereas negation absorption comes as a completely new and stipulated mechanism, the NCC enforces structural identities, just as most HPSG principles do (such as the identity of HEAD values in the HEAD FEATURE PRINCIPLE).

The NCC also makes the correct predictions for non-clausal contexts such as the *albo* construction exemplified in (39) above. There, we get at most one negation, even if there are several n-words:[18]

(51) a. Poślubię albo tę dziewczynę z Poznania, albo żadną inną dziewczynę z żadnego miasta.
I will marry either [this girl].ACC from Poznan or [no other girl].ACC from no city.
'I will either marry this girl from Poznan or no other girl from any city.'

b. Piotr opowiada albo Marii coś o swojej pracy albo nikomu w ogóle nic.
Peter tells either Mary something about his work or nobody at all nothing
'Peter tells either something about his work to Mary or nothing to anybody at all.'

Instances of *double negation* are only permitted if the negations are not part of the same head projection. Thus, in (52) the higher verb *może* is in the scope of a single negation and so is the lower verb *znać* within its own projection.[19]

(52) 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 NCC enforces the NC reading of Polish n-words. The NCC does, however, not enforce the presence of a pre-verbal *nie* in verbal contexts, i.e., we cannot exclude sentence (53).

(53) * Nikt przyszedł.
Nobody came

[18]Just as would be expected from the data in (39), the NP *no other girl* in (51a) could also appear in genitive, i.e. as *żadnej innej dziewczyny*.
[19]For a more detailed discussion of this example, see Section 4.2.3.

Figure 6: Structure for the ungrammatical sentence **Nikt przyszedł.*:

$$\begin{bmatrix} \text{EXCONT} & \boxed{4}\ \neg\exists x[\mathsf{human}'(x) \wedge \mathsf{come}'(x)] \\ \text{INCONT} & \boxed{1}\ \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}, \boxed{3}, \boxed{3a}, \boxed{4}, \boxed{5}, \boxed{5a} \right\rangle \end{bmatrix} \qquad \begin{array}{l} \& \boxed{5} \triangleleft \beta \\ \& \boxed{3} \triangleleft \gamma \\ \& \boxed{1} \triangleleft \delta \end{array}$$

$$\underset{\textit{Nikt}}{\begin{bmatrix} \text{EXCONT} & \boxed{5}\ \exists x[\gamma \wedge \delta] \\ \text{INCONT} & \boxed{3}\ \mathsf{human}'(x) \\ \text{PARTS} & \left\langle x, \boxed{3}, \boxed{3a}\mathsf{human}', \boxed{4}\neg\beta, \boxed{5}, \boxed{5a}[\gamma \wedge \delta] \right\rangle \end{bmatrix}} \qquad \underset{\textit{przyszedł}}{\begin{bmatrix} \text{EXCONT} & \boxed{4} \\ \text{INCONT} & \boxed{1}\ \mathsf{come}'(x) \\ \text{PARTS} & \left\langle x, \boxed{1}, \boxed{1a}\mathsf{come}' \right\rangle \end{bmatrix}}$$

Thus, a second language-specific principle must guarantee the presence of *nie* in negated clauses. Following Haegeman and Zanuttini (1991) we call it the NEG CRITERION.

(54) The NEG CRITERION (NegC):

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

Corblin and Tovena (2001) argue that such a principle is just an instance of a general tendency in languages to express negation "pre-verbally", their *Neg first Principle*. This principle has several possible forms, and our formulation of the NegC for Polish would fall within the parameter range of this Principle.[20] Our approach differs from the approach in Corblin and Tovena (2001) in that we refer to the explicit presence of a negation operator in the semantic representation, whereas the formulation in Corblin and Tovena (2001) seems to be compatible with a syntactic feature NEG. Similarly, the Neg Criterion in Haegeman and Zanuttini (1991) is clearly syntactic, as it also refers to particular syntactic nodes and is formulated as an instance of head-specifier agreement.

For illustration of the NegC, consider the small syntactic tree in Figure 6. The only element of the overall PARTS list that can satisfy the EContP is the expression $\boxed{4}$ that contains the negation contributed by the n-word. According to the SP, the EXCONT value of the head is identical to that of the phrase. Thus, the verb *przyszedł* has an EXCONT value which contains a negation, and the negation has in turn scope over its INCONT value. The NegC requires, then, that the negation also be part of the verb's PARTS list, which it is not. The structure in Figure 6 is, thus, excluded by the NegC.

Conclusion Our analysis of NC in Polish shows how token identity can be used in the derivation of logical representations to arrive at a natural account of concord phenomena. Whereas our analysis is similar in spirit to GB analyses in the school of Haegeman and Zanuttini (1991), we do not need any stipulated mechanisms such as *negation factorization*. Instead, we can simply use the major analytical tool of HPSG, token identity. Cross-linguistic research of the kind exemplified in Corblin and Tovena (2001) has also shown that in any conceivable analysis of the negation phenomena in question, two independent and language specific principles are necessary. The NCC and the NegC can be seen as fully formalized instances of these principles.

With the examples discussed so far, it could be shown that the obligatory presence of pre-verbal *nie* is not an argument for excluding an inherent negative character of n-words when LRS is used. In the following section, we extend our fragment to include opaque verbs and modal verbs to show that the data in (42) can be equally well handled within the present account and, thus, do not constitute an empirical argument against the negativity of n-words, but rather an argument for the inadequacy of the combinatorial techniques used in LF-Ty2 and other non-discontinuous frameworks.

4.2.3 Opaque Cases

The basic problem that was illustrated in the examples in (42) is that there is some semantic material that may intervene between the negation and the existential quantifier which is contributed by an n-word. This material includes modal

[20] In fact their data suggest that their analysis of Romanian could be taken directly over to Polish.

operators and intensional verbs. Given the architecture of LRS, it is, however, clear that nothing in the lexical entry of the n-word in (48) requires that the existential quantifier be in the *immediate* scope of the negation. After all, the very idea of a discontinuous framework like LRS is that it be appropriate for interweaving the semantic contribution of an element with that of other elements. The same had already been illustrated for the semantic contribution of the verb. The negated verb *nie przyszedł* in the small example in Figure 5 contributes a negation and the expression $\mathsf{come}'(x)$, but the existential quantifier of the n-word intervenes between them. As we are going to demonstrate now, this technique can be generalized to an intensional fragment.

The lexical entry for *nikt* in (48) was very simplistic. To account for the data with intensional and modal verbs, we must include a treatment of intensionality. We follow the Montagovian analysis of Hendriks (1993), but use Ty2 just as in Groenendijk and Stokhof (1982). Following their example, we write $@$ for the distinguished world variable $v_{s,0}$. Montague's $\hat{}\phi$ becomes $\lambda @.\phi$ (see Gallin (1975)). Taking intensionality into account, the specification in (48) of the lexical entry of *nikt* changes into that given in (55).

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

$$\begin{bmatrix} \textit{word} & \\ \textsc{phon} & \langle \textit{nikt} \rangle \\ \textsc{synsem loc cont} & \begin{bmatrix} \textit{lrs} & \\ \textsc{excont} & \boxed{5}\ \exists x[\gamma \wedge \delta] \\ \textsc{incont} & \boxed{3}\ \mathsf{human}'_{@}(x_{@}) \\ \textsc{parts} & \left\langle @, x, x_{@}, \boxed{3}, \boxed{3a}\mathsf{human}'_{@}, \boxed{3b}\mathsf{human}', \boxed{4}\neg\beta, \boxed{5}, \boxed{5a}[\gamma \wedge \delta] \right\rangle \end{bmatrix} \end{bmatrix} \begin{array}{l} \&\ \boxed{5} \triangleleft \beta \\ \&\ \boxed{3} \triangleleft \gamma \end{array}$$

$$\&\ \mathbf{A}\boxed{7}\left(\left(\boxed{7}\left[\textsc{var}\ \textit{variable}\right] \&\ \boxed{7} \triangleleft \delta\ \&\ \texttt{is-free}(x,\boxed{7})\right) \Rightarrow \boxed{7}\left[\textit{existential}\right]\right)$$

The major changes are the following: As mentioned above, we write $@$ for the distinguished world variable $v_{s,0}$. Individual constants are still assumed to be of type e, but individual variables are taken to be intensional, i.e., x is now of type se. Therefore, we write $x(@)$ (which corresponds to $\check{}x$ in IL) if an individual variable occurs in the argument slot of a predicate. The constant human' also takes an additional world argument. To keep the notation shorter, we sometimes write the world argument as subscript. Thus, instead of $\mathsf{human}'(v_{0,s}, x(v_{s,0}))$, we write $\mathsf{human}'(@, x(@))$, or even $\mathsf{human}'_{@}(x_{@})$. The rest of the AVM is as presented in (48), including the two constraints at the side of the AVM.

We have further conjoined a restriction to the main AVM box of the lexical entry that determines which kinds of elements may intervene between the quantifier ($\exists x[\gamma \wedge \delta]$) and the occurrences of x within the scope of this quantifier. It is evident from the data that we have considered so far that no operator may intervene here, with the exception of other existential operators. This means that the n-word always takes narrow scope with respect to all other operators in the clause.

While this restriction to narrow scope is a property of the quantificational part of the n-word, it does not apply to the negation part of its semantic contribution. For illustration, consider sentence (42a) again, repeated in (56). In Polish, this sentence only has the *de dicto* reading. This follows from the intervention restriction in the lexical entry of the n-word: In the *de re* reading, the variable x is free in the scope of the intensional operator $\lambda @$.

(56) Janek nie szuka żadnego jednorożca.
Janek NM seeks no unicorn

a. *de dicto*: $\lambda @.\neg\mathsf{seek}'_{@}(j, \lambda @\lambda P.\exists x[\mathsf{unicorn}'_{@}(x_{@}) \wedge P_{@}(x_{@})])$

b. *de re*: \$ $\lambda @.\neg\exists x[\mathsf{unicorn}'_{@}(x_{@}) \wedge \mathsf{seek}'_{@}(j, \lambda @\lambda P.P_{@}(x_{@}))]$

In order to see how the *de dicto* reading is derived, consider the description of the negated form of the opaque verb *nie szuka* (*NM seeks*), shown in (57). To guide the reader through it, we have added brief explanations to the right of the subexpression restrictions that appear below the AVM.

(57) Description of the word *nie szuka* (*NM seeks*):

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nie szuka} \rangle \\ \text{SYNS LOC CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{8}\, P_{@}(y_{@}) \\ \text{PARTS} & \langle @, x, y, P, y_{@} \\ & \boxed{1}\mathsf{seek}'_{@}(x, \lambda @ \lambda P.\epsilon), \boxed{1a}\mathsf{seek}'_{@}(\lambda @ \lambda P.\epsilon), \boxed{1b}\mathsf{seek}'_{@}, \boxed{1c}\mathsf{seek}', \\ & \boxed{0}\lambda @.\zeta, \boxed{2}\neg\alpha, \boxed{7}\lambda @ \lambda P.\epsilon, \boxed{7a}\lambda P.\epsilon, \boxed{8}P_{@}(y_{@}), \boxed{8a}P_{@} \rangle \end{bmatrix} \end{bmatrix}$$

& $\boxed{1} \triangleleft \alpha$ (seek' is in the scope of $\neg$)
& $\boxed{2} \triangleleft \zeta$ ($\neg$ is in the scope of the EXCONT's $\lambda @$)
& $\boxed{8} \triangleleft \epsilon$ ($P_{@}(y)$ is in the scope of seek')

It should be noted that, contrary to the intuitions that we have appealed to so far, the INCONT value of *nie szuka* does not contain the semantic constant seek', but rather the scopally lowest element contributed by the word, i.e., the expression $P_{@}(y_{@})$. This interpretation of the INCONT value is fully consistent with the examples discussed so far and also with the SP as given in (30).[21]

From the lexical entry in (55) and the description in (57) we can construct the VP *nie szuka niczego* (*NM seeks nothing*). The *lrs* of this VP is just like that of the verb *nie szuka*, but also contains the expressions of the PARTS list of the n-word, and the additional constraint that the INCONT value of the verb, the expression $P_{@}(y_{@})$, be in the scope of the n-word, i.e., be a subexpression of δ. Combining this VP with the non-quantificational subject *Janek* does not introduce any new restrictions on the reading.

We can, now, see that the *de dicto* reading given in (56a) is the only reading that satisfies all the constraints imposed by the lexical specifications, the SP and the other principles (IContP, EContP, NCC and NegC). The narrow scope of the n-word with respect to the opaque operator $\lambda @$ is enforced by the lexical entry of the n-word. However, its having scope over the expression $P_{@}(y_{@})$ follows from the SP. While the n-word and the negated verb both contribute a negation, the identity of these negations is required by the NCC in (50). In its description, the negated verb already specifies that its negation must have scope over the semantic constant seek'. Thus, a narrow scope for the negation is excluded.

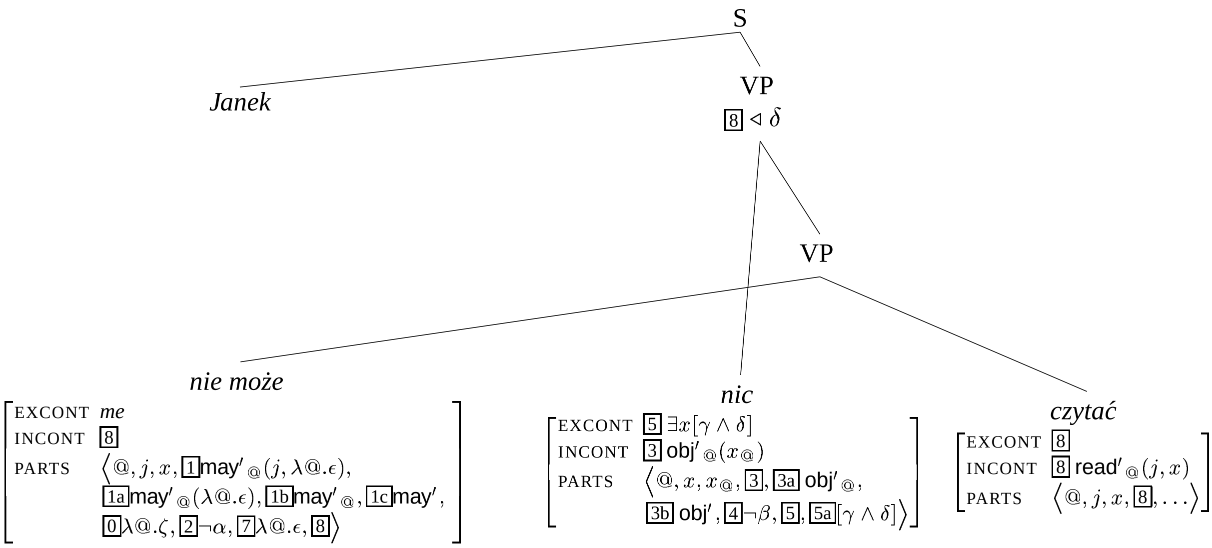
A similar account can be given for constructions with modal verbs, exemplified in (42b), and repeated for convenience in (58).

(58) Janek nie może nic czytać.
Janek NM can nothing read 'Janek is not allowed to read anything.'
$\neg\mathsf{may}'_{@}(j, \lambda @.\exists x[\mathsf{object}'_{(}x_{@}) \wedge \mathsf{read}'_{@}(j, x_{@})])$

For the modal verbs, we assume a syntactic analysis along the lines of Przepiórkowski (1999a). Modals are "argument raisers" (Hinrichs and Nakazawa, 1989), but raising does not cross *nie*-prefixed verbs. To get the word order right, we assume an order domain approach of the kind proposed for German (Kathol, 2000; Reape, 1992; Richter and Sailer, 1995; Richter, 1997) and for Slavonic (Penn, 1999).

[21] In Section 5, we will present a cleaner architecture, in which we will distinguish between the basic semantic contribution of a lexical element and its scopally lowest semantic contribution.

Figure 7: The analysis of (58):



(59) Description of the word *nie może* (*NM can*):

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \textit{nie może} \rangle \\ \text{SYNS LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SUBCAT} & \boxed{9} \oplus \left\langle \begin{bmatrix} \text{L} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HD} & \textit{verb} \\ \text{SBC} & \boxed{9} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{EXCONT} & \boxed{8} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{lrs} & \\ \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{8} \\ \text{PARTS} & \left\langle @, y, \boxed{1}\mathsf{may}'_{@}(y, \lambda @.\epsilon), \boxed{1a}\mathsf{may}'_{@}(\lambda @.\epsilon), \boxed{1b}\mathsf{may}'_{@}, \boxed{1c}\mathsf{may}', \right. \\ & \left. \boxed{0}\lambda @.\zeta, \boxed{2}\neg\alpha, \boxed{7}\lambda @.\epsilon, \boxed{8}, \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \begin{array}{l} \& \ \boxed{1} \triangleleft \alpha \\ \& \ \boxed{2} \triangleleft \zeta \\ \& \ \boxed{8} \triangleleft \epsilon \end{array}$$

Just as we saw for the verb *nie szuka* in (57), the INCONT value of a modal is not the modal operator, but rather the scopally lowest element in the semantic contribution of the verb. In this case, we assume it to be the EXCONT value of its verbal complement, given as $\boxed{8}$.[22]

Figure 7 shows the structure of the sentence in (58). The fact that there is only a *de dicto* reading for the existential quantifier is, again, a consequence of the restriction in the lexical entry of the n-word. The fact there is only one negation is a result of the NCC. The wide scope of the negation with respect to the modal is a consequence of the *nie* prefixation on the verb.

Before closing this section, we have yet to demonstrate that the double negation reading of examples with more than one negated verb does not pose problems for the analysis. Consider again example (34), repeated below:

(60) 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 EXCONT value of the verbal complement *nie znać Marii* is the expression $\lambda @.\neg\mathsf{know}'_{@}(t, m)$. That the negation contributed by the verb *nie znać* must be a part of this expression is required by the prefixation as indicated

[22]This analysis is reminiscent of the CONTENT raising analysis given for semantically empty (vacuous) verbs such as *to* in Pollard and Sag (1994) or Pollard and Yoo (1998).

in all the descriptions of *nie* prefixed verbs given so far. The negation is realized within the syntactic head projection of the verb. The NegC is satisfied for the lower VP, as the negation that appears in the EXCONT value is also part of the PARTS list. The same is valid for the NCC, as there is only one negation in the lower VP.

At the clause level, the INCONT value of the modal verb is identical to the EXCONT value of the lower VP, i.e., it is the expression $\lambda @.\neg \mathsf{know}'_{@}(t, m)$. The overall EXCONT value introduces another negation, such that in the logical form of the sentence, there are two negations. Nonetheless the NCC is satisfied, because at the upper VP, only one negation has scope over the INCONT value of *nie może*, because the lower negation is a component of the INCONT value of *nie może*.

Our analysis of intensional contexts has shown the interaction of discontinuity and potential identities of logical contributions of different words. In these contexts, the negation contributed by the negation prefix *nie* and an n-word can be identified, while the quantifier contributed by the n-word is realized in a position in the logical expression that is non-adjacent to the negation symbol. At the same time, additional restrictions may constrain the kind of operators that may or may not intervene between the negation and the existential quantifier of the n-word. Since we are at liberty to use the powerful description language of RSRL to impose constraints on logical expressions, these constraints can in principle be very complex. Of course, we hope that in practice they follow a rigid and simple pattern that does not interfere with an efficient computational treatment.

4.2.4 A Brief Evaluation of the Analysis

The major feature of our LRS analysis of NC in Polish is the fact that it easily avoids mechanisms such as *negative absorption*, which have previously been employed to eliminate multiple, superfluous copies of negation introduced by several lexical elements. The schema in (61) illustrates what *negative absorption* does in such analyses. The symbol $\Longrightarrow$ stands for a transformation operation that takes a logical formula that follows the pattern stated to its left and yields a formula that obeys the pattern to its right.

(61) $\forall x_1 \neg \forall x_2 \ldots \neg \forall x_n \neg \phi \Longrightarrow \forall x_1 \forall x_2 \ldots \forall x_n \neg \phi$

An extra-logical transformation of logical formulae is clearly not desirable in any grammar framework. LRS replaces it with the attractive alternative of a constraint language that can talk about logical representations and offers the option of interpreting the constraints of different lexical entries as talking about the same pieces of a logical language. In HPSG this is in fact a well-established and central method of grammatical description and is known under the name of *structure sharing* or structural *identities*. "Structure sharing" has been a device at the very heart of grammar design of HPSG from the beginnings of the framework, and as such it is a familiar device whose application to the present problem provides a systematic solution that agrees with the rest of the architecture of grammar. Since structural identities may occur anywhere in linguistic structures where they are not excluded by principles of grammar or the ontology of objects,[23] NC phenomena in languages can even be regarded the unmarked case that occurs if not explicitly prevented by some language-specific principles.

The principles that we introduced, NEGATION COMPLEXITY CONSTRAINT and NEG CRITERION, are adaptations to LRS of standard principles in this area of research. They are simply re-phrasing commonly accepted empirical generalizations in an appropriately precise technical language. Their realization in LRS shows that LRS is apt to integrate typical grammatical generalizations at the syntax-semantics interface quite easily.

When we compare the LRS analysis of NC in Polish to alternative analyses in the semantic frameworks that we have considered in this paper (Ty2U, Ty2U^P, and LF-Ty2), we observe a number of advantages. Since LRS does not work with underspecified semantic representations, it does not have the descriptive problems of Ty2U; it does not have the technical problems connected to the use of meta-variables and the limited range of restrictions that can be put on the form of logical representations in Ty2U and Ty2U^P. Finally, from a descriptive point of view, an analysis in LF-Ty2 of the same empirical domain is much more complex (Richter and Sailer, 1999b), and, as we have argued, cannot come as close to the intuitions about the inherent negativity of n-words that guided the present analysis. The opaque contexts of Section 4.2.3 all but exclude a similar LF-Ty2 analysis, since it does not have the discontinuous semantic representations of lexical units that LRS can appeal to.

[23] As determined by the signature. See Section 2.1.1 for the relevant underlying assumptions about grammars of RSRL.

4.3 Summary

In contrast to traditional systems of underspecified semantics like Ty2U, LRS does not introduce an additional level of underspecification to grammars. Instead, LRS stands in the tradition of old-fashioned logical form semantics: In each unembedded expression of a natural language—expressions with illocutionary force—there is an attribute whose value is a logical form. This logical form is a true and fully specific representation of the meaning of the expression in a logical language. For our introduction to LRS in the present paper, we chose Ty2 for its familiarity among linguists.

LRS deliberately exploits the fact that a description in the grammar specification languages of RSRL may describe many structurally different objects. In the case of our logical representations of Ty2 (or any other appropriate class of formal languages), the objects in the denotation of our grammar specification languages are logical expressions. Thus we have a natural place for underspecification in the grammar, namely the constraint language that talks about the logical representations. However, underspecification on that level does not lead to underspecified denotations of semantic representations, and it does not involve indirect representations like the meta-variables of Ty2U and Ty2U^P.

In effect, the descriptive and technical problems of denotationally underspecified or indirect systems are avoided, while desirable aspects of underspecification are lifted to the level of description of logical representations; the possibility for discontinuity in semantic representations is retained; many insights into possible advantageous structurings of logical representations gained in connection with Ty2U are captured in the structure of *lrs* objects; and new descriptive possibilities such as those involving structural identities are added to the inventory of analytical devices.

5 Local Semantics

In Section 4.2.3 we mentioned that we cannot maintain the original intuition behind the function of the attribute INCONT—expressing the genuine semantic contribution of a sign—when it comes to opaque and modal contexts. In the present section, we will modify the feature geometrical position of semantics within our HPSG grammar by modifying the signature. While this will not touch upon the LRS mechanism itself, it will lead to a more structured approach to semantic representation.

One of the major empirically motivated changes in the feature geometry of signs that took place between (Pollard and Sag, 1987) and (Pollard and Sag, 1994) was the introduction of a more refined theory of selection. While the valence features in (Pollard and Sag, 1987) contain lists of signs, only *synsem* objects can be selected in (Pollard and Sag, 1994). As a consequence, the phonological shape and the syntactic constituent structure of the selected element is made invisible for the selector. However, the semantic structure of the selected element is still fully available for it in the architecture of Pollard and Sag (1994). This is odd, as selection is immune with respect to certain aspects of the logical form such as quantifier scope relations. In this section, we will introduce a distinction within the semantic representation that is parallel to the category/structure distinction in syntax. Doing so we achieve a more restricted theory of selection. Along the way, our modification also retrieves a location in all words where we specify their genuine semantic contribution, including opaque and intensional verbs like *seek* and *may*.

The major motivation for the design of LRS, and of all related systems with discontinuous semantic representations (Egg, 1998; Egg and Erk, 2002; Copestake et al., 2000; Richter and Sailer, 1999c) stems from the empirical domain of scope ambiguities. Since these aspects of the semantics are not relevant for selection, we will change the appropriateness conditions of the sort *sign* in the way indicated in (62). We introduce a new attribute LOGICAL-FORM (LF). The value of this attribute is an LRS representation, i.e., an object of sort *lrs* as defined in (24).

(62) Appropriateness conditions for the sort *sign*:

sign	PHON	*phonology*
	SYNSEM	*synsem*
	L(OGICAL-)F(ORM)	*lrs*

With this change, there is no need to have an LRS representation as the CONTENT value inside *synsem* objects. In (Pollard and Sag, 1994), the CONTENT fulfills a number of functions that we actually lost in the previous sections. First, it contains the index of nominals. This index is necessary for HPSG's BINDING THEORY. Second, all theories of linking in HPSG, such as (Koenig and Davis, 2001), rely on the basic semantic contribution of a lexical element. As became clear in the analysis of opaque verbs in Section 4.2.3, the INCONT value does not contain the basic semantic contribution of a lexical element in the general case. Thus, there is no single attribute whose value would uniquely identify the semantic term that must be available for a theory to express the linking between syntactic and semantic arguments.

It is important to note that index information and linking information, in contrast to scope information, are needed inside *synsem* objects in the theory of Pollard and Sag (1994). Therefore, it makes perfect sense to retain this information as part of the CONTENT value. In (63) we state the new appropriateness conditions for the sort *content*.

(63) Appropriateness condition for the sort *content*:

content	INDEX	*extended-index*
	MAIN	*term*

Since we do not intend to develop a detailed theory of local semantics, we limit our discussion to this very simple form of the local content. The MAIN value inside the CONTENT value of a lexical sign is now used to identify the basic semantic constant that is contributed by this particular sign.

For the values of the attribute INDEX we introduce a new sort, *extended-index*, whose own appropriate attributes and attribute values are shown in (64).

(64) Appropriateness conditions for the sort *extended-index*:

extended-index	PHI	*index*
	VAR	*term*

Our new kind of index has two attributes. First, an attribute PHI which contains exactly the structure of objects of sort *index* in (Pollard and Sag, 1994), which are basically the traditional ϕ-features, number, gender and person. The need for a more structured INDEX value was originally mentioned in (Soehn, 2003) and (Soehn and Sailer, 2003), where the attribute PHI was introduced. Second, the value of the attribute VAR contains the term of the semantic representation language that corresponds to the index in the logical forms. Typically, this term is an individual variable. Note that in contrast to Pollard and Sag (1994), but following Soehn (2003), we have an *index* on all parts of speech, including verbs. For verbs we assume that the INDEX VAR value is an event variable, while the INDEX PHI value of verbs does not play any role in the grammar and remains underspecified in the lexical entry of a verb.

The new architecture is illustrated with the lexical entries of *Peter* and the verb *read* in (65) and (66), respectively.

(65) Sketch of the lexical entry of *Peter*:

$$\left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \textit{Peter} \rangle \\ \text{SYNSEM} & \left[\begin{array}{ll} \text{LOC CONTENT} & \left[\begin{array}{ll} \textit{content} & \\ \text{INDEX} & \left[\begin{array}{ll} \text{PHI} & \left[\begin{array}{ll} \text{GEN} & \textit{masc} \\ \text{NUM} & \textit{sg} \\ \text{PERS} & \textit{3rd} \end{array}\right] \\ \text{VAR} & \boxed{1}\, p \end{array}\right] \\ \text{MAIN} & \boxed{1} \end{array}\right] \end{array}\right] \\ \text{LF} & \left[\begin{array}{ll} \textit{lrs} & \\ \text{EXCONT} & \textit{me} \\ \text{INCONT} & \boxed{1} \\ \text{PARTS} & \left\langle \boxed{1} \right\rangle \end{array}\right] \end{array}\right]$$

(66) Sketch of the lexical entry of *read*:

$$\left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \textit{read} \rangle \\ \text{SYNS} & \left[\begin{array}{ll} \text{LOC} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \text{HEAD} & \textit{verb} \\ \text{SUBCAT} & \left\langle \text{NP}\left[\text{INDEX VAR } x\right], \text{NP}\left[\text{INDEX VAR } y\right] \right\rangle \end{array}\right] \\ \text{CONT} & \left[\begin{array}{ll} \text{INDEX} & \left[\text{VAR } e\right] \\ \text{MAIN} & \mathsf{read}' \end{array}\right] \end{array}\right] \end{array}\right] \\ \text{LF} & \left[\begin{array}{ll} \text{EXCONT} & \boxed{1} \\ \text{INCONT} & \mathsf{read}'(e,x,y) \\ \text{PARTS} & \left\langle e, x, y, \exists e.\phi, ((\mathsf{read}' e)y)x, (\mathsf{read}' e)y, \mathsf{read}' e, \mathsf{read}' \right\rangle \end{array}\right] \end{array}\right] \begin{array}{l} \&\ \exists e.\phi \triangleleft \boxed{1} \\ \&\ \mathsf{read}'(e,x,y) \triangleleft \phi \end{array}$$

It should be noticed that we do not have to change the LRS principles that we introduced in Section 4.1 (the IContP, the EContP, and the SP), because these principles only concern the semantic combinatorics and constrain the possible scopings. They do not mention the basic semantic contribution of lexical elements. This seems to be a desirable fact, because it shows that the general phrasal aspects of semantics are independent of the particular lexical semantic contribution.

Matters change, however, when we consider the role of particular lexical elements. The principles that we have introduced in our analysis of Polish in Section 4.2 are concerned with the interaction of the semantic contribution of lexical items. Intuitively, the NegC should express the idea that if the core lexical semantic contribution of a verb is in the scope of a negation in a clause then this negation must be part of the semantic contribution of the verb. As discussed in Section 4.2.3, however, the attributes EXCONT and INCONT do not always allow us to refer to this core lexical contribution. With the new architecture of local content, we can finally provide a more adequate formulation of the relevant principles. In fact, it suffices to refer to the MAIN value instead of the INCONT value in the NEGATION COMPLEXITY CONSTRAINT and in the NEG CRITERION. Additionally, it is necessary to change the description of negated verbs in such a way that the negation operator introduced by the prefix *nie* must have scope over the MAIN value of the verb, not over its INCONT value.

Making these changes accounts for the data involving opaque verbs described in Section 4.2.3. Note that the verb *seek* now receives its intuitively plausible MAIN value inside the local content, while its scopally lowest semantic contribution, the term $P_{@}(y_{@})$, is its INCONT value.

In (67) we describe the relevant parts of the negated verb *nie szuka* (*NM seek*) according to our new architecture.

(67) Description of the word *nie szuka* (*NM seeks*):

$$\begin{bmatrix} \textit{word} \\ \text{PHON } \langle \textit{nie szuka} \rangle \\ \text{SYNS LOC CONT} \begin{bmatrix} \text{INDEX} & \begin{bmatrix} \text{VAR } e_{@} \end{bmatrix} \\ \text{MAIN} & \boxed{1d}\ \mathsf{seek}' \end{bmatrix} \\ \text{LF} \begin{bmatrix} \textit{lrs} \\ \text{EXCONT} \ \textit{me} \\ \text{INCONT} \ \boxed{8}\ P(@, y) \\ \text{PARTS} \ \langle @, x, y, e, P, e_{@}, y_{@} \\ \quad \boxed{1}\mathsf{seek}'(@, e_{@}, x, \lambda @ \lambda P.\epsilon), \boxed{1a}\mathsf{seek}'(@, e_{@}, \lambda @ \lambda P.\epsilon), \boxed{1b}\mathsf{seek}'(@, e_{@}), \boxed{1c}\mathsf{seek}'(@), \boxed{1d}\mathsf{seek}', \\ \quad \boxed{0}\lambda @.\zeta, \boxed{2}\neg\alpha, \boxed{7}\lambda @ \lambda P.\epsilon, \boxed{7a}\lambda P.\epsilon, \boxed{8}P(@, y_{@}), \boxed{8a}P(@) \rangle \end{bmatrix} \end{bmatrix}$$

$\& \ \boxed{1d} \triangleleft \alpha$ (seek' is in the scope of $\neg$)
$\& \ \boxed{2} \triangleleft \zeta$ ($\neg$ is in the scope of the excont $\lambda @$)
$\& \ \boxed{8} \triangleleft \epsilon$ ($P_{@}(y_{@})$ is in the scope of seek')

Analogously, for modals, we can differentiate between their basic semantic contribution, which is the modal operator, and the scopally lowest element in their logical form. In the case of modals, we have seen that an NP that is realized syntactically as an argument of the modal can still be in the scope of the modal. Therefore, we must assume that the INCONT value of the modal is identical to the INCONT value of its verbal complement.

Due to our changes in the signature, however, the INCONT value of the complement is not contained in the *synsem* object that the modal verb selects. Instead, it is only present in the phrase where the modal and its verbal complement are combined. Therefore, the identity of the INCONT values of the head and the non-head daughter must be enforced by a principle of grammar rather than in a lexical entry.

If we inspect the constellations in which this identity of INCONT values occurs, we encounter a systematic pattern: INCONT identity of the head and the non-head is required whenever the syntactic head combines with a constituent whose complements are raised to the SUBCAT list of the head. Argument raising is a lexical property of a number of verbs and adjectives, and INCONT identities are a predictable consequence of this property.

Thus, parallel to argument raising, we speak of *content raising* in the case of INCONT identity of head and non-head. The parallelism lies in the fact that some basic properties of a selected element become properties of the selector: They are complements in the case of argument raising, and the INCONT-value in the case of content raising.

In (68) we state the CONTENT RAISING PRINCIPLE, which enforces the described INCONT identities.

(68) CONTENT RAISING PRINCIPLE:

$$\mathbf{A}\,\boxed{1}\left(\begin{bmatrix}\text{H-DTR} & \left[\text{SYNS LOC CAT SUBCAT } \langle\ldots\rangle \oplus \boxed{1} \oplus \langle\ldots\rangle\right]\\ \text{N-DTR} & \left[\text{SYNS LOC CAT SUBCAT } \boxed{1}\ \textit{nelist}\right]\end{bmatrix}\Rightarrow \mathbf{E}\,\boxed{2}\left(\begin{bmatrix}\text{LF} & \left[\text{INCONT } \boxed{2}\right]\\ \text{N-DTR} & \left[\text{LF } \left[\text{INCONT } \boxed{2}\right]\right]\end{bmatrix}\right)\right)$$

The CONTENT RAISING PRINCIPLE requires INCONT identities, whenever the complements of the non-head ($\boxed{1}$) have been raised to the SUBCAT list of the head. In the consequent of the principle, we only mention the INCONT values of the phrase and of the non-head daughter. It should be noticed that the INCONT value of the head daughter is necessarily identical to that of the entire phrase, because of Clause 2 of the SEMANTICS PRINCIPLE in (30).

Given these prerequisites, we can reconsider the structure of sentence (58), repeated in (69).

(69) Janek nie może nic czytać.
Janek NM can nothing read

$\lambda @.\exists e'\neg \mathsf{may}'(@, e'_{@}, \lambda @.\exists e \exists x[\mathsf{object}'(@, x_{@}) \wedge \mathsf{read}'(@, e_{@}, j, x_{@})])$

In (70) we describe the negated modal verb *nie może* (*NM may*) according to our new architecture. In Figure 8 we see the resulting analysis.

(70) Description of the word *nie może* (*NM can*):

$$\begin{bmatrix}\textit{word}\\ \text{PHON } \langle \textit{nie może}\rangle\\ \text{SYNS LOC} \begin{bmatrix}\text{CAT} & \left[\text{SUBCAT } \boxed{9} \oplus \left\langle \left[\text{L } \left[\text{CAT } \begin{bmatrix}\text{HD} & \textit{verb}\\ \text{SBC} & \boxed{9}\end{bmatrix}\right]\right]\right\rangle\right]\\ \text{CONT} & \begin{bmatrix}\text{INDEX} & \left[\text{VAR } e'_{@}\right]\\ \text{MAIN} & \mathsf{may}'\end{bmatrix}\end{bmatrix}\\ \text{LF} \begin{bmatrix}\textit{lrs}\\ \text{EXCONT} & \textit{me}\\ \text{INCONT} & \boxed{8}\\ \text{PARTS} & \left\langle @, e', e'_{@}, \boxed{1}\mathsf{may}'(@, e'_{@}, \epsilon), \boxed{1a}\mathsf{may}'(@, e'_{@}), \boxed{1b}\mathsf{may}'(@), \boxed{1c}\mathsf{may}', \right.\\ & \left. \boxed{0}\lambda @.\zeta, \boxed{0a}\exists e'.\phi', \boxed{2}\neg\alpha, \boxed{7}\epsilon, \boxed{8}, \right\rangle\end{bmatrix}\end{bmatrix} \begin{array}{l}\&\ \boxed{1} \triangleleft \alpha\\ \&\ \boxed{2} \triangleleft \zeta\\ \&\ \boxed{8} \triangleleft \epsilon\end{array}$$

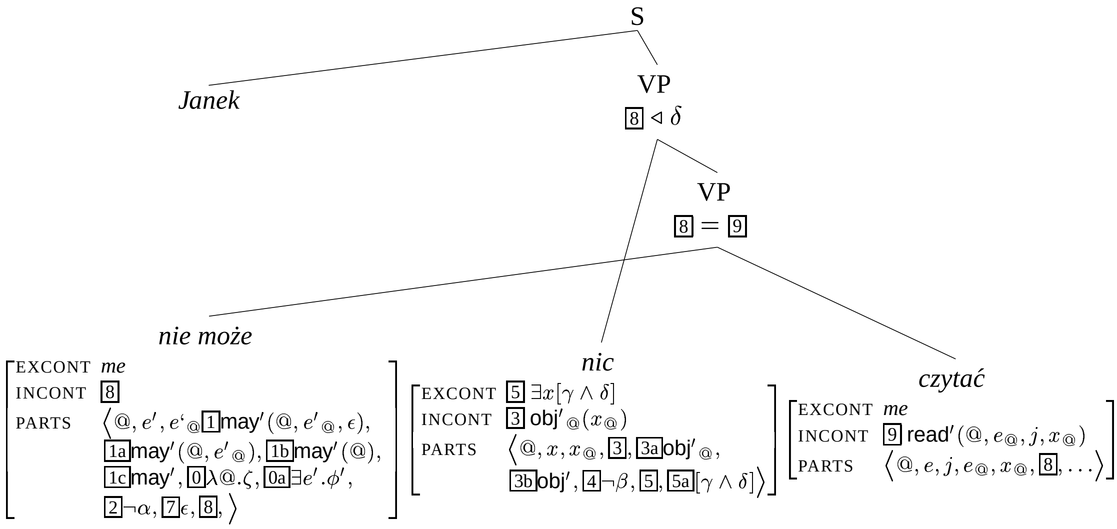
In this section, we introduced a distinction between CONTENT and LF which parallels the distinction between CATEGORY and DAUGHTERS in Pollard and Sag (1994). The new distinction allows us to use LRS as sketched in Section 4.1 without any changes, and still accommodate the original HPSG BINDING THEORY and approaches to linking in HPSG. Besides these technical or integrative advantages, the new architecture results in a more constrained theory of selection. Nonetheless this more restricted theory of selection is fully compatible with the empirical findings.

6 Conclusions and Outlook

With the introduction of a distinction between the local semantics under CONTENT and the nonlocal semantics under LF in the previous section, we have made LRS compatible with the traditional BINDING THEORY of HPSG and with theories of linking. Moreover, we have been able to locate the genuine semantic contribution of lexical items in the feature geometry, after we lost such a location during the extension of LRS to cover intensional contexts. As a result, we have developed LRS into a theory of semantics that is compatible with all components of the theory of grammar presented in (Pollard and Sag, 1994), and adds a complete semantic meta-theory to it. We have compared LRS to Ty2U, Ty2U^{P}, and LF-Ty2, and we have argued at length that it combines the descriptively necessary properties of semantic theories with technical achievements of the systems we examined.

An obvious next step is to put LRS to the test and extend the empirical coverage of the grammars in which it is applied to see how well it scales up. On this side of developments, we need to find out whether LRS can support theoretically satisfying and well-founded analyses of further phenomena.

Figure 8: The analysis of (69):



A second very important aspect of LRS is its computational implementation. In cooperation with Gerald Penn we are working on an implementation of LRS as a module of the TRALE grammar development environment. For this project, it turned out to be necessary to design a specification language for LRS that greatly simplifies the separation of logical representations into feature value descriptions of three attributes (EXCONT, INCONT and PARTS) in *lrs* objects. Our specification language for the computational implementation now actually uses a single decorated term to express the same information as an entire description of *lrs* objects. We believe that this is a notation that is also interesting for theoretical LRS grammar specifications as in this paper. However, in our paper we have avoided the highly compressed notation of our LRS implementation language for several reasons. The most important reason is that a notation that is closer to an RSRL constraint language is more transparent in an introduction to the topic, since it remains clearer on which level of the grammatical architecture constraints and denotations are located. Moreover, it is much easier to demonstrate the connections between Ty2U and LRS when using a unified notation.

Currently we have a prototype of LRS together with an implementation of a fragment of German in TRALE. Just as in the examples of the present paper, the fragment of German employs Ty2. We are also planning to implement a small grammar of Polish designed to test the soundness of our analysis of Polish NC in a computational environment. Unfortunately, there are no conclusive results on the efficiency of the implementation yet, and it is too early to predict the behavior of the system with larger grammars.

Finally, as we have emphasized from the outset, LRS is designed to be a meta-theory that can be applied to various classes of semantic representation languages. The first class of logical object languages we used came from Ty2, since Ty2 is one of the standards in natural language semantics. Once we have a stable implementation of LRS in TRALE, encouraging results about the performance of the system, and experience with larger natural language fragments with LRS semantics, we intend to turn from Ty2 to other, more recent formal languages for the representation of meaning in natural languages.

References

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

Błaszczak, Joanna (1998). Some Reflexions on the Interaction between Indefinites and the Clause: The Case of Indefinite Pronouns. In M. Giger and B. Wiemer (Eds.), *Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV)*, Volume 1, pp. 55–65. München.

Błaszczak, Joanna (1999). *Investigations into the Interaction between the Indefinites and Negation*. Ph. D. thesis, Humboldt Universität Berlin. Published in: Akademieverlag, Berlin, Studia grammatica 51, 2001.

Bos, Johan (1996). Predicate Logic Unplugged. In P. Dekker and M. Stokhof (Eds.), *Proceedings of the Tenth Amsterdam Colloquium*, pp. 133–143. ILLC/Department of Philosophy, University of Amsterdam.

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.

Bouma, Gosse (2003). Verb Clusters and the Scope of Adjuncts in Dutch. In P. A. M. Seuren and G. Kempen (Eds.), *Verb Constructions in German and Dutch*. Amsterdam and Philadelphia: Benjamins.

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`.

Copestake, Flickinger, Dan, Sag, Ivan, and Pollard, Carl (2000, March). Minimal Recursion Semantics. An Introduction. Manuscript, Stanford University.

Corblin, Francis (1994). Multiple Negation Processing. Research Paper HCRC/RP-62, Human Communication Research Center.

Corblin, Francis (1995). Compositionality and Complexity in Multiple Negation. *IGPL Bulletin 3*(3–2), 449–473.

Corblin, Francis and Tovena, Lucia M. (2001). On the Multiple Expression of Negation in Romance. In Y. D'Hulst, J. Rooryck, and J. Schroten (Eds.), *Romance Languages and Linguistic Theory 1999*, pp. 87–115. John Benjamin, Amsterdam.

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 Erk, Katrin (2002). A Compositional Account of VP Ellipsis. In F. Van Eynde, L. Hellan, and D. Beermann (Eds.), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pp. 162–179. Stanford: CSLI Publications.

Egg, Markus, Koller, Alexander, and Niehren, Joachim (2001). The Constraint Language for Lambda Structures. *Journal of Logic, Language and Information 10*(4), 457–485.

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.

Giannakidou, Anastasia (1998). *Polarity Sensitivity as Nonveridical Dependency*. John Benjamins.

Ginzburg, Jonathan and Sag, Ivan A. (2000). *Interrogative Investigations. The Form, Meaning, and Use of English Interrogatives*. CSLI Publications.

Groenendijk, Jeroen and Stokhof, Martin (1982). Semantic Analysis of wh-Complements. *Linguistics and Philosophy 5*, 175 –233.

Haegeman, Liliane and Zanuttini, Raffaella (1991). Negative Heads and the Neg Criterion. *The Linguistsic Review 8*, 233–251.

Halvorsen, Per-Kristian (1995). Situation Semantics and Semantic Interpretation in Constraint-Based Grammars. In M. Dalrymple, R. M. Kaplan, J. T. I. Maxwell, and A. Zaenen (Eds.), *Formal Issues in Lexical-Functional Grammar*, pp. 293–309.

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

Hinrichs, Erhard W. and Nakazawa, Tsuneko (1989). Flipped Out: Aux in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistics Society*, Chicago, IL, pp. 193–202.

Kasper, Robert (1997). Semantics of Recursive Modification. Unpublished Manuscript. The Ohio State University.

Kathol, Andreas (2000). *Linear Syntax*. Oxford University Press.

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.

Koenig, Jean-Pierre and Davis, Anthony (2001). Sublexical Modality and the Structure of Lexical Semantic Representations. *Linguistics and Philosophy 24*, 71–124.

Kupść, Anna (2000). *An HPSG Grammar of Polish Clitics.* Ph. D. thesis, Polish Academy of Sciences and Université Paris 7.

Kupść, Anna and Przepiórkowski, Adam (1998). Morphological Aspects of Verbal Negation in Polish. Paper delivered at the Second International Conference on Formal Description of Slavic Languages, Potsdam, November 20–22, 1997. To appear in the proceedings.

Ladusaw, William (1980). *Polarity Sensitivity as Inherent Scope relations*. Garland Press.

Linebarger, Marcia Christine (1980). *The Grammar of Negative Polarity*. Ph. D. thesis, MIT. cited after the reproduction by the Indiana University Linguistics Club, Indiana, 1981.

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.

Penn, Gerald (1999). Linearization and WH-Extraction in HPSG: Evidence from Serbo-Croatian. In R. D. Borsley and A. Przepiórkowski (Eds.), *Slavic in HPSG*, pp. 247–282. CSLI Publications.

Pinkal, Manfred (1996). Radical Underspecification. In P. Dekker and M. Stokhof (Eds.), *Proceedings of the Tenth Amsterdam Colloquium*, pp. 587 – 606. ILLC/Department of Philosophy, University of Amsterdam.

Pinkal, Manfred (1999). On Semantic Underspecification. In H. Bunt and R. Muskens (Eds.), *Computing Meaning*, Studies in Linguistics and Philosophy, pp. 33–55. Kluwer Academic Publishers.

Pollard, Carl and Sag, Ivan A. (1987). *Information-Based Syntax and Semantics. Volume 1: Fundamentals*. CSLI Lecture Notes 13.

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.

Pollard, Carl J. and Yoo, Eun Jung (1998). A Unified Theory of Scope for Quantifiers and *wh*-Phrases. *Journal of Linguistics 34*, 415–445.

Przepiórkowski, Adam (1999a). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph. D. thesis, Universität Tübingen.

Przepiórkowski, Adam (1999b). Negative Polarity Questions and Italian Negative Concord. In V. Kordoni (Ed.), *Tübingen Studies in HPSG*, Arbeitspapiere des SFB 340, Nr. 132, Volume 2, pp. 353–400. Universität Tübingen.

Przepiórkowski, Adam and Kupść, Anna (1997a). Negative Concord in Polish. Technical Report 828, Institute for Computer Science, Polish Academy of Sciences.

Przepiórkowski, Adam and Kupść, Anna (1997b). Verbal Negation and Complex Predicate Formation in Polish. In *Proceedings of the 1997 Conference of the Texas Linguistics Society on the Syntax and Semantics of Predication*, Number 38 in Texas Linguistic Forum, pp. 247–261.

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.

Reape, Mike (1992). *A Formal Theory of Word Order: A Case Study of West Germanic*. Ph. D. thesis, University of Edinburgh.

Reyle, Uwe (1993). Dealing with Ambiguities by Underspecification: Construction, Representation and Deduction. *Journal of Semantics 10*(2), 123–179.

Richter, Frank (1997). Die Satzstruktur des Deutschen und die Behandlung langer Abhängigkeiten in einer Linearisierungsgrammatik. Formale Grundlagen und Implementierung in einem HPSG-Fragment. In E. Hinrichs, W. D. Meurers, F. Richter, M. Sailer, and H. Winhart (Eds.), *Ein HPSG-Fragment des Deutschen. Teil 1: Theorie*, Arbeitspapiere des SFB 340, Nr. 95, pp. 13–187. Universität Tübingen.

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

Richter, Frank and Sailer, Manfred (1995). Remarks on Linearization. Reflections on the Treatment of LP-Rules in HPSG in a Typed Feature Logic. Magisterarbeit, Seminar für Sprachwissenschaft, Universität Tübingen.

Richter, Frank and Sailer, Manfred (1996). Syntax für eine unterspezifizierte Semantik: PP-Anbindung in einem deutschen HPSG-Fragment. In S. Mehl, A. Mertens, and M. Schulz (Eds.), *Präpositionalsemantik und PP-Anbindung*, pp. 39–47. Gerhard-Mercator-Universität Duisburg.

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 (2001a). On the Left Periphery of German Finite Sentences. In W. D. Meurers and T. Kiss (Eds.), *Constraint-Based Approaches to Germanic Syntax*, pp. 257–300. CSLI Publications.

Richter, Frank and Sailer, Manfred (2001b). Polish Negation and Lexical Resource Semantics. In G.-J. Kruijff, L. S. Moss, and R. T. Oehrle (Eds.), *Proceedings FG-MOL 2001*, Number 53 in Electronic Notes in Theoretical Computer Science. Elsevier Science. To appear.

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.

Riehemann, Susanne (1997). Idiomatic Constructions in HPSG. Paper presented at the *Fourth International Conference on Head-Driven Phrase Structure Grammar*, Ithaca. `http://doors.stanford.edu/~sr/idioms.ps`.

Riehemann, Susanne Z. (2001). *A Constructional Appraoch to Idiomatic Expressions and Word Formation*. Ph. D. thesis, Stanford University.

Sag, Ivan A. (1997). English Relative Clause Constructions. *Journal of Linguistics 33*, 431–483.

Sailer, Manfred (1997). Adjunkte in einer Linearisierungsgrammatik: Syntax, Semantik und Wortstellung. In E. Hinrichs, W. D. Meurers, F. Richter, M. Sailer, and H. Winhart (Eds.), *Ein HPSG-Fragment des Deutschen. Teil 1: Theorie,* Arbeitspapiere des SFB 340, Nr. 95, pp. 249–318. Universität Tübingen.

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

Soehn, Jan-Philipp (2003). Von Geisterhand zu Potte gekommen. Eine HPSG-Analyse von PPs mit unikaler Komponente. Magisterarbeit, Universität Tübingen, Seminar für Sprachwissenschaft.

Soehn, Jan-Philipp and Sailer, Manfred (2003). At First Blush on Tenterhooks. About Selectional Restrictions Imposed by Nonheads. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner (Eds.), *Proceedings of Formal Grammar 2003*, pp. 149–161.

Witkoś, Jacek (1996). On NegP and the Structure of the Polish Clause. *papers and Studies in Contrastive Linguistics 31*, 65–96.

Zwarts, Frans (1981). Negatief polaire uitdrukkingen. *GLOT 4*(1), 35–132.