Chapter 12

Closer to the Truth: A New Model Theory for HPSG

Frank Richter
Eberhard Karls Universität Tübingen
Tübingen, Germany
`fr@sfs.uni-tuebingen.de`

12.1 Introduction

HPSG is a model theoretic grammar framework in which a grammar is formulated as a pair consisting of (a) a signature which generates a space of possible structures and (b) a set of grammar principles which single out the well-formed structures among them. There are three proposals of how to precisely define the denotation of grammars within this general setting. Each proposal is accompanied by its own meta-theory of the ontological nature of the structures in the denotation of the grammar and their relationship to empirically observable phenomena. I will show that all three model theories face serious, if not fatal, problems: One of them makes very idiosyncratic fundamental assumptions about the nature of linguistic theories which many linguists might not share; the other two fail to capture the concepts they were designed to make mathematically precise. I will propose an alternative model theory which takes into account the shape of actual grammars and fixes the shortcomings of its predecessors.

12.2 The Plot

HPSG is an attractive candidate for studying a model theoretic linguistic framework. It has a history of over 20 years, many HPSG grammars of different languages have been written, and there are mathematically precise proposals about the denotation of HPSG grammars. Thus it is possible to take actual grammar specifications written by linguists and investigate the classes of structures the grammars denote according to the different model theories.

Here I want to take advantage of this fortunate situation to address the following questions:

1. Do the models of HPSG grammars meet the apparent intentions of the linguists who write them? And if they do not, how can we repair the problem(s) as conservatively as possible?

2. Are the structures in the denotation of the grammars actually compatible with the meta-theories of the meaning of grammars formulated within the HPSG framework?

The paper proceeds in three steps. Section 12.3 reviews problems with models of typical grammars (irrespective of the choice of meta-theory) and suggests universal restrictions on the form of HPSG grammars to amend them. Section 12.4 presupposes these amendments and investigates the models which the existing three meta-theories postulate. In response to the shortcomings we find, Section 12.5 proposes a new definition of the meaning of HPSG grammars, together with a meta-theory of the relationship between the set of structures denoted by an HPSG grammar and empirical linguistic phenomena. In the final section I conclude with a few remarks on the relationship of the new proposal to its predecessors.

For space reasons, I will concentrate on a rather informal discussion of the problems and their solutions. The presentation of the mathematical details is left for a different occasion.

12.3 Imprecise Grammars

Instead of taking a realistic grammar of a natural language as my object of study, I approach the questions of Section 12.2 with a very simple toy grammar which is built in such a way that it reflects crucial properties which all actual HPSG grammars in the literature share. This simplification helps to keep our modeling structures at a manageable (i.e., readable) size. Crucially, for our toy grammar below it will be obvious which structures form its intended denotation, and we can easily investigate whether the logical formalism supports the apparent expectations of the linguist.

12.3.1 An Example

An HPSG grammar consists of (a) a signature, Σ, declaring a sort hierarchy, attribute appropriateness conditions, and a set of relations and their arity, and (b) a set of logical statements, θ,usually called the *principles of grammar*. The grammar $\langle \Sigma_1, \theta_1 \rangle$ in (7) and (8) is a particularly simple example which, however, is structured like a typical linguistic grammar.

A most general sort, *top*, is the supersort of all other sort symbols in the sort hierarchy. The attributes PHON (for phonology) and CAT (syntactic category) are appropriate to all signs, with values *list* and *cat*, respectively. Attribute appropriateness is inherited by more specific sorts, in this case *word* and *phrase*, with the possibility of subsorts adding further appropriate attributes. Here the sort *phrase* also bears the attributes H_DTR (head daughter) and NH_DTR (non-head daughter) for the syntactic tree structure. Another important feature of the present signature is the attribute SUBCAT, appropriate to *cat*. SUBCAT will be used for the selection of syntactic arguments. Finally, the signature introduces a relation symbol for a ternary relation, `append`.

(7) The signature Σ_1:

```
top
    sign      PHON     list
              CAT      cat
        phrase    H_DTR     sign
                  NH_DTR    sign
        word
    list
        nelist    FIRST   top
                  REST    list
        elist
    cat     HEAD      head
            SUBCAT    list
    head
        verb
        noun
    phonstring
        uther
        walks
append/3
```

The signature Σ_1 together with the theory θ_1 predicts exactly three well-formed signs: The words *Uther* and *walks* and the phrase *Uther walks*. The idea is that *Uther* and *walks* are not only words in our grammar, they may also occur as complete independent utterances, e.g. in exclamations and elliptical statements. θ_1 incorporates important HPSG principles: A WORD PRINCIPLE specifies the well-formed words, a (trivial) IMMEDIATE DOMINANCE (ID) PRINCIPLE specifies admissible phrase structures, a HEAD FEATURE PRINCIPLE makes category information travel up syntactic head projections, and a CONSTITUENT ORDER PRINCIPLE regulates word order. The last principle fixes the intended meaning of the relation symbol `append`.

(8) The theory θ_1:

a. WORD PRINCIPLE:

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

b. ID PRINCIPLE:

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

c. HEAD FEATURE PRINCIPLE:

$$[\textit{phrase}] \longrightarrow \begin{bmatrix} \text{CAT HEAD } \boxed{1} \\ \text{H\_DTR CAT HEAD } \boxed{1} \end{bmatrix}$$

d. CONSTITUENT ORDER PRINCIPLE:

$$[\textit{phrase}] \longrightarrow \left(\begin{bmatrix} \text{PHON} & \boxed{3} \\ \text{H\_DTR PHON} & \boxed{2} \\ \text{NH\_DTR PHON} & \boxed{1} \end{bmatrix} \wedge \texttt{append}(\boxed{1},\boxed{2},\boxed{3}) \right)$$

e. APPEND PRINCIPLE:

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

Models only contain objects labeled with maximally specific sorts (sorts without any proper subsorts in the sort hierarchy). For each appropriate attribute, there is one outgoing arc which points to an object labeled with an appropriate maximally specific sort. Informally, HPSG grammars denote a class of structures comprising all structures licensed by the signature such that all nodes in these structures also obey the well-formedness requirements imposed by the theory. In other words, the denotation of the grammar comprises at least one copy of each possible well-formed structure. Such 'complete' models are called *exhaustive models*.

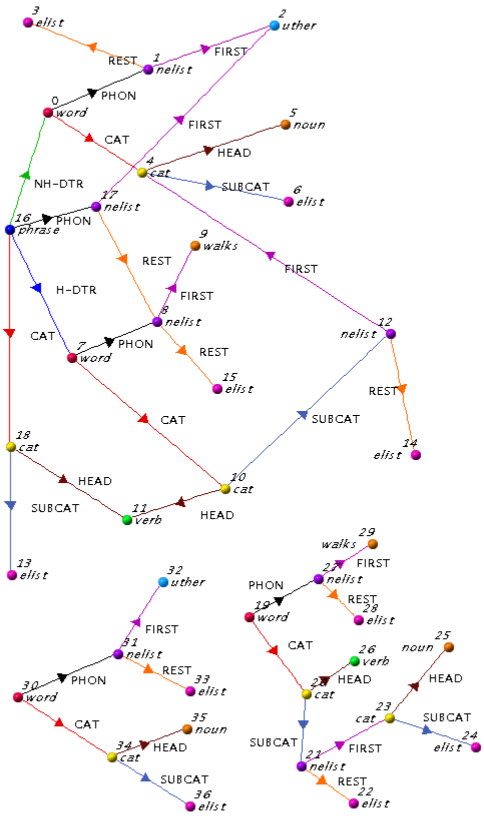
Which structures do linguists expect to find in the denotation of grammar $\langle \Sigma_1, \theta_1 \rangle$? Fig. 12.1 shows the most likely candidate (omitting the relation). The configuration with the phrasal root node 16 represents the sentence *Uther walks*; the configurations with root nodes 30 and 19 represent the words *Uther* and *walks*.

Upon reflection it is not difficult to see that these are by far not the only configurations licensed by our grammar. Three kinds of problems can be readily distinguished, which I will call the *intensionality of lists*, *twin structures*, and *stranded structures*.

The *intensionality of lists* is a side effect of the particular feature logical encoding of lists standardly adopted in HPSG. Consider the structure for the word *walks* under node 19 above. It contains three distinct *elist* objects (22, 24, 28) at the end of the PHON and SUBCAT lists of the verb and at the end of the SUBCAT list of its selected argument. Nothing in the grammar prevents any two or even all three *elist* objects from being the same object. This way we get five possible configurations for the word *walks* which the linguist presumably never intended to distinguish. We should clearly treat this ambiguity as an accident of encoding and get rid of it.

Twin structures are structures with more than one root node. For example, nothing would prevent the HEAD arc originating at the subcategorized object 23 in the word *walks* from pointing to the object 35 of the word *Uther* instead of to the object 25. The *noun* object 35 would then belong to the word *walks* and to the word *Uther*. No restrictions of the grammar would be violated, but what em-

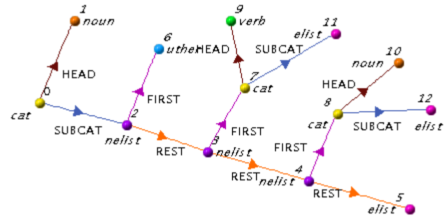
Figure 12.1: The intended $\langle \Sigma_1, \theta_1 \rangle$ model



pirical phenomenon should correspond to linguistic structure belonging to two (or even more) independent utterances? It seems obvious to me that this kind of configuration is not intended by linguists, and it should not occur in the intended models. In this paper I will not elaborate on the causes of the problem and on the full range of possible solutions. It will disappear as a side effect of the solution to the third problem of our grammar, stranded structures.

Stranded structures constitute the most serious one of the three types of problems with the grammar $\langle \Sigma_1, \theta_1 \rangle$. Stranded structures are typically structures which are 'smaller' than utterances. As an immediate consequence, they tend to be inaccessible to empirical observation. A trivial example is a configuration which looks just like the configuration under the *cat* object 34 of *Uther* in Fig. 12.1, the only difference being that there is no arc pointing to the *cat* object: It is stranded and inaccessible to empirical observation, since it is not connected to a phonological value. While some of the stranded structures in the denotation of grammars are isomorphic to structures which occur in observable linguistic signs (such as the one just described), *stranded monster structures* are of a shape which prevents them from being possible substructures of well-formed linguistic signs. Fig. 12.2 shows such a monster in the denotation of $\langle \Sigma_1, \theta_1 \rangle$.

Figure 12.2: A stranded monster structure in a $\langle \Sigma_1, \theta_1 \rangle$ model



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

The monster in Fig. 12.2 is a nominal *cat* object whose SUBCAT list contains the phonetic string *Uther* and selects a verb and a noun. Although no such category exists in a word in the denotation of our grammar, it exists as a stranded structure because the constraints that prevent its existence in words all operate at the sign level. It is immediately clear that our grammar denotes infinitely many stranded monster structures. Even worse, the architecture of signs in HPSG and the standard grammar principles guarantee the existence of infinite classes of stranded monster structures in realistic grammars.

Contrary to first appearances, there is no simple remedy for this problem. Consider a brute force restriction which states that only configurations with root nodes of sort *word* and *phrase* may populate the linguistically relevant models, configurations which are empirically accessible through their phonology. However, there are phrases which require a licensing environment. In HPSG this environment may in fact contribute crucial structural restrictions, and its absence leads to absurd phrasal structures. Slashed constituents – phrases which contain an extraction site for a constituent without their corresponding filler – are a straightforward example. Their semantics will partly depend on the extracted constituent as recorded in the SLASH set. According to HPSG signatures, configurations in SLASH are smaller than signs (they are of sort *local*). Moreover, there are hardly any well-formedness restrictions on these *local* configurations as long as the extracted constituent is not realized as a sign in the syntactic tree. Therefore the configurations under *local* objects in the SLASH set of a slashed constituent without its complete licensing environment are usually not configurations which may actually occur in signs according to the grammar principles. A slashed constituent without its embedding matrix environment might thus have an arbitrary and even impossible semantics, due to the unrestricted *local* configuration in SLASH and its contribution to the meaning of the constituent. This means that monster structures are back, and this time they even have a phonology and make empirically false predictions.

The grammars in the HPSG literature are not precise enough for their models to match the intentions of linguists. Independent of the choice of model theory they denote structures that their authors do not intend to predict. As the considerations about slashed constituents show, this is not a problem of the model theories. It is preferable to solve it by amending the grammars.

12.3.2 Normal Form Grammars

What we saw in the previous section was a weakness of the linguistic theory rather than of the logical formalism. Stranded structures are often inaccessible to empirical observation and should not be predicted. In grammars with interesting coverage stranded structures also materialize as phrasal stranded monster structures. These have a phonology, which means that they should be observable, but their internal structure prevents them from occurring as part of an actual utterance.

Appropriate extensions of the linguistic theory eliminate the spurious structures and can simply be added to most HPSG grammars. The extensions consist of general assumptions about the signature and of a number of logical statements to be included among the grammar principles.

The first move is to single out utterances from other types of signs as the only ones that are immediately empirically accessible. Every kind of linguistic structure is ultimately part of an utterance. Since no linguistic structure can simultaneously belong to two utterances, twin structures are ruled out. A minor technical amendment concerns lists: For their encoding we fix a unique structure that excludes spurious ambiguities that stem from multiple *elist* objects. In sum, I add to each HPSG grammar

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

A grammar which incorporates these restrictions will be called a *normal form grammar*. The signature of the normal form grammar derived from the grammar $\langle\Sigma_1,\theta_1\rangle$ is shown in (10). The hierarchy of signs distinguishes between unembedded signs (*u_sign*) and embedded signs (*e_sign*), a distinction which is inherited by words and phrases. Syntactic daughters are always embedded signs. The specification in the signature of the EMBEDDED value *u_sign* for each object ensures that every object in an interpretation is tied to an unembedded sign. The dots under *list* stand for all declarations under *list* in (7), including `append`.

(10) Normal form extension Σ_2 of signature Σ_1:

```
top   EMBEDDED  u_sign
   sign   PHON  list
          CAT   cat
      e_sign
         e_word
         e_phrase
      u_sign
         u_word
         u_phrase
      word
         e_word
         u_word
      phrase  H_DTR   e_sign
              NH_DTR  e_sign
         e_phrase
         u_phrase
   list
...
component/2
```

(11) shows the logical statements which must be added to the theory θ_1 in (8) to obtain the corresponding normal form grammar $\langle\Sigma_2,\theta_2\rangle$. The new theory, θ_2, incorporates all principles from θ_1 in (8), adding four new restrictions on admissible models. For each of the new principles the corresponding formulation in (9) is indicated. The relation `component` is defined with respect to all attributes $\mathcal{A}$ in the signature. (11i) states that each pair of nodes x and y in a configuration is in the `component` relation iff a sequence of attributes leads from y to x.

(11) Normal form extension θ_2 of theory θ_1:[1]

f. (3c) U-SIGN COMPONENT CONDITION:

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

g. (3d) UNIQUE U-SIGN CONDITION:

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

h. (3e) UNIQUE EMPTY LIST CONDITION:

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

i. COMPONENT PRINCIPLE:

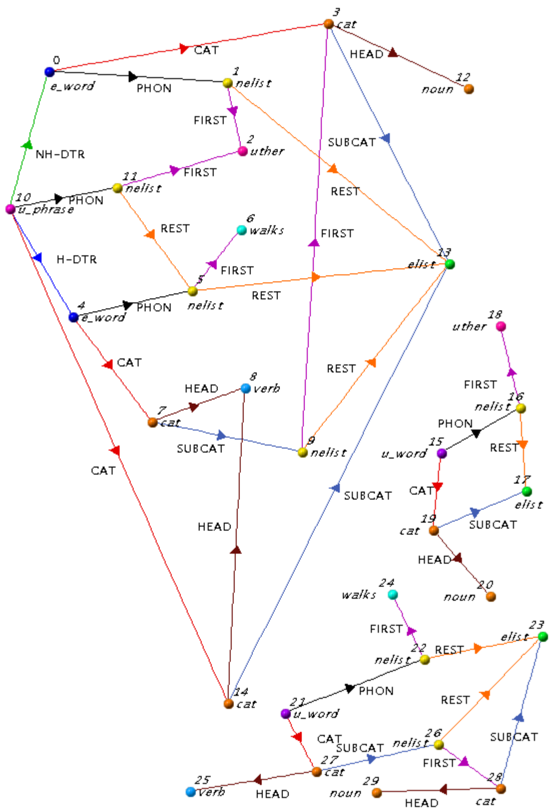
$$\forall\boxed{1}\forall\boxed{2}\left(\begin{array}{l}\mathtt{component}(\boxed{1},\boxed{2}) \leftrightarrow \\ \left(\begin{array}{l}\boxed{1}=\boxed{2}\ \vee \\ \bigvee\limits_{\alpha\in\mathcal{A}} \exists\boxed{3}\left(\boxed{2}[\alpha\ \boxed{3}] \wedge \mathtt{component}(\boxed{1},\boxed{3})\right)\end{array}\right)\end{array}\right)$$

The effect of normalizing the grammar $\langle\Sigma_1,\theta_1\rangle$ can be inspected in Fig. 12.3. For readability I systematically omit the attribute EMBEDDED, which points from each node to the unique *u_sign* node to which the node belongs. For example, each node in the configuration with

[1]The logical expressions are RSRL descriptions (Richter, 2004). '$\forall$' is not the first order universal quantifier.

the *u_phrase* 10 – representing the sentence *Uther walks* – has an outgoing EMBEDDED arc pointing to 10. The reader may want to verify that there are no other possible configurations in the denotation of the grammar. It should also be noted that the independent words *Uther* (under *u_word* node 15) and *walks* (under *u_word* node 21) are no longer isomorphic to the occurrences of these words in the sentence, because they are now marked as unembedded.

Figure 12.3: An exhaustive $\langle\Sigma_2,\theta_2\rangle$ model, systematically omitting the attribute EMBEDDED for readability (see the explanation in the text)



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

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

12.4 Problems in Previous Model Theories

On the basis of the notion of *normal form HPSG grammars* I can now investigate the previous mathematical characterizations of the meaning of HPSG grammars. These are (1) Pollard and Sag's original theory of linguistic utterance types modeled by abstract feature structures (Pollard and Sag 1994), (2) Pollard's theory of mathematical idealizations of utterance tokens (Pollard 1999), and (3) King's theory of exhaustive models containing sets of possible utterance tokens (King 1999). In order to make sure that all three logical formalisms can easily be compared and are comprehensive enough for a full formalization of HPSG grammars of the kind introduced by Pollard and Sag (1994), I use them in their variants defined in (Richter, 2004), which expresses them in terms of Relational Speciate Re-entrant Language (RSRL).

12.4.1 Informal Overview

The formalization of the model theory of (1) and (2) fails to produce models that agree with their respective meta-theories of the structures in their grammar models. In essence, the problem is that both (1) and (2) intend to capture the idea that for each isomorphism class of well-formed utterances in a language, we find exactly one structure in the denotation of the grammar which models the members of the isomorphism class. For example, take a realization of the utterance *I am sitting in a 370 year old house in Engadin.* The intention of the model theory of (1) is to have exactly one abstract feature structure in the denotation a grammar of English which models – or stands for the utterance type of – the utterance token. Similarly, the intention of the model theory of (2) is to have exactly one mathematical idealization of the isomorphism class of tokens of the given sentence in the denotation of the grammar. However, this intention is not borne out in either formalism. Their models are defined in such a way that we necessarily find a large number of modeling structures for the given sentence in the denotation of a correct grammar of English. Subsection 12.4.2 sketches the properties of the formalisms which are responsible for this result.

The problem with (3) is not of a technical nature, it comes from the meta-theory itself. King postulates that the intended model of a grammar is an *exhaustive model* like the one shown in Fig. 12.3 for the grammar $\langle\Sigma_2,\theta_2\rangle$. According to King, the exhaustive model of a language that the linguist aims for does not contain utterance types or mathematical idealizations of utterance tokens. Instead it contains the utterance tokens of the language themselves. Since we cannot know how many tokens of a given utterance there have been and will be in the world, we never know how many isomorphic copies of each utterance token the intended model contains. The definition of exhaustive models permits an arbitrary number of isomorphic copies of each possible configuration, all that is required is the presence of at least one representative of each. From the definition we only know that the class of exhaustive models of a grammar comprises, among many others, the particular exhaustive model which, for each utterance, contains the right number of tokens (if the grammar is correct). However, since there will be grammatical utterances of a language which have never occurred and will never occur, this is not yet the full story. As exhaustive models (by definition) contain at least one copy of each potential grammatical utterance in the language, the intended exhaustive model must also comprise *possible* (as opposed to actual) utterance tokens, at least for those well-formed utterances of a language which never occur. This means that the configurations in exhaustive models are *potential utterance tokens*. These potential utterance tokens are a dubious concept if tokens are supposed to be actual occurrences of a linguistic form. In light of this problem, King's model theory has been unacceptable to some linguists.

12.4.2 Details

In this section I substantiate my claim that the model theories based on abstract feature structures by Pollard and Sag (1994) and on mathematical idealizations of linguistic utterance tokens by Pollard (1999) do not achieve what their meta-theories call for. Henceforth I refer to these two theories as AFS and MI, respectively.

Let us first consider AFS. The underlying idea is that the denotation of a grammar is a set of *relational abstract feature structures* as determined by an *admission relation*. Each abstract feature structure in the set of relational abstract feature structures admitted by a grammar is a unique representative of exactly one utterance type of the natural language which the grammar is supposed to capture. This means that there is a one-to-one correspondence between the utterance types of the natural language and the abstract feature structures which the grammar admits. A grammar can then be falsified by showing either that there is no feature structure admitted by the grammar which corresponds to a particular utterance type of the language or that the grammar admits an abstract feature structure which does not correspond to any grammatical utterance type in the language.

Relational abstract feature structures consist of four sets: A *basis set*, β, which provides the basic syntactic material; a *re-entrancy relation*, ρ, which is an equivalence relation that can be understood as an abstract representation of the nodes in connected configurations; a *label function*, λ, which assigns species to the abstract nodes; and a *relation extension*, symbolized below as ξ, which represents the tuples of abstract nodes which are in the relations of a grammar.

How these four components of a relational abstract feature structure conspire to produce a representation of the utterance type *Uther* from Fig. 12.3 can be seen in (12).[2] The symbol ε stands for the empty path, i.e., an empty sequence of attributes. The basis set, β_U, contains all attribute paths which can be created by following sequences of arcs from 15. The re-entrancy relation, ρ_U, enumerates all possibilities of getting to the same node by a pair of attribute paths; and the label function, λ_U, assigns the correct species to each attribute path.

[2]For expository purposes I pretend that the attribute EMBEDDED is not in the grammar. See footnote 3 for further remarks on this simplification.

(12) $\mathbb{A}_{\text{Uther}} = \langle \beta_U, \rho_U, \lambda_U, \xi_U \rangle$ with

$$
\begin{aligned}
\beta_U &= \left\{ \begin{array}{l} \varepsilon, \text{PHON}, \text{PHON REST}, \text{PHON FIRST}, \\ \text{CAT}, \text{CAT SUBCAT}, \text{CAT HEAD} \end{array} \right\}, \\
\rho_U &= \left\{ \begin{array}{l} \langle \varepsilon, \varepsilon \rangle, \langle \text{PHON}, \text{PHON} \rangle, \langle \text{CAT}, \text{CAT} \rangle, \\ \langle \text{PHON FIRST}, \text{PHON FIRST} \rangle, \\ \langle \text{PHON REST}, \text{PHON REST} \rangle, \\ \langle \text{PHON REST}, \text{CAT SUBCAT} \rangle, \\ \langle \text{CAT SUBCAT}, \text{PHON REST} \rangle, \\ \langle \text{CAT SUBCAT}, \text{CAT SUBCAT} \rangle, \\ \langle \text{CAT HEAD}, \text{CAT HEAD} \rangle \end{array} \right\}, \\
\lambda_U &= \left\{ \begin{array}{l} \langle \varepsilon, \mathit{u\_word} \rangle, \langle \text{PHON}, \mathit{nelist} \rangle, \\ \langle \text{PHON REST}, \mathit{elist} \rangle, \\ \langle \text{CAT SUBCAT}, \mathit{elist} \rangle, \\ \langle \text{PHON FIRST}, \mathit{uther} \rangle, \langle \text{CAT}, \mathit{cat} \rangle, \\ \langle \text{CAT HEAD}, \mathit{noun} \rangle \end{array} \right\}, \\
\xi_U &= \left\{ \begin{array}{l} \langle \mathtt{append}, \text{PHON}, \text{PHON REST}, \text{PHON} \rangle, \\ \langle \mathtt{append}, \text{PHON REST}, \text{PHON}, \text{PHON} \rangle, \\ \langle \mathtt{append}, \text{PHON}, \text{CAT SUBCAT}, \text{PHON} \rangle, \\ \langle \mathtt{append}, \text{CAT SUBCAT}, \text{PHON}, \text{PHON} \rangle \end{array} \right\} \\
&\cup \left\{ \langle \mathtt{append}, \pi_1, \pi_2, \pi_3 \rangle \;\middle|\; \begin{array}{l} \pi_1, \pi_2, \pi_3 \in \\ \left\{ \begin{array}{l} \text{PHON REST}, \\ \text{CAT SUBCAT} \end{array} \right\} \end{array} \right\} \\
&\cup \left\{ \langle \mathtt{component}, \pi_1, \pi_2 \rangle \;\middle|\; \begin{array}{l} \pi_1 \in \beta_U, \pi_2 \in \beta_U, \& \\ \pi_1 = \pi_2 \text{ or} \\ \pi_2 \text{is a prefix of } \pi_1 \end{array} \right\}
\end{aligned}
$$

Note that the set theoretical definition of abstract feature structures guarantees that every abstract feature structure isomorphic to another one is identical with it.

Figure 12.4: The utterance type *Uther* and its reducts, without relations and the EMBEDDED attribute

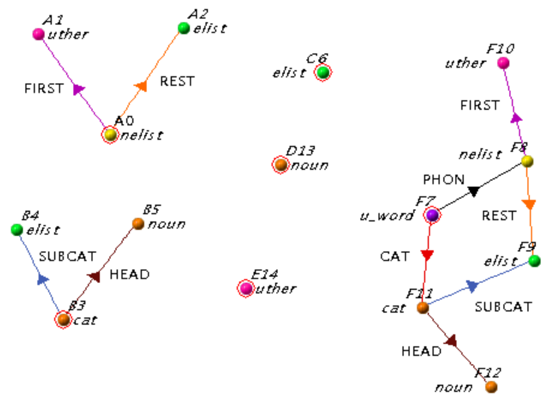


Fig. 12.4 repeats the *Uther* configuration from Fig. 12.3 and adds a few more configurations. They are all rooted at a distinguished node (marked by a circle). The significance of the new configurations is the fact that the set of abstract feature structures admitted by our grammar does not only contain the abstract feature structure corresponding to the *Uther* configuration under $F7$ (beside the two corresponding to *walks* and *Uther walks*).

Since the abstract feature structure for *Uther* is in the set, it also contains abstract feature structures corresponding to the configurations under $A0$, $B3$, $C6$, $D13$ and $E14$.

The reason for this is to be found in the definition of relational abstract feature structures and the ensuing admission relation based on the traditional satisfaction relation for feature structures, and it is an artifact of the construction. Intuitively, this is what happens: Abstract feature structures lack an internal recursive structure. Since the admission relation must ensure that the entire abstract feature structure including all of its abstract nodes satisfies the set of principles of a grammar, an auxiliary notion of reducts provides the necessary recursion. The idea is that a relational abstract feature structure is admitted by a theory if and only if the feature structure itself and all its reducts satisfy the theory. But that means that not only the relational abstract feature structure but also all of its reducts are in the set of abstract feature structures admitted by the theory.

The definition of reducts is straightforward. Any attribute path in the basis set may be followed to get to an abstract node in the feature structure. At the end of each path we find a new abstract root node of a reduct. This can best be seen by considering the corresponding pictures of configurations in Fig. 12.4 again. The configuration under $A0$ corresponds to the PHON reduct of the *Uther* configuration; the configuration under $B3$ corresponds to the CAT reduct of the *Uther* configuration; $C6$ to the PHON REST and CAT SUBCAT reduct; and analogously for the two remaining atomic configurations. (13) contains an example of the reducts depicted in Fig. 12.4, an abstract feature structure corresponding to the configuration with root node $E14$. The reducts can be obtained either by abstraction from the configurations in Fig. 12.4 or directly from $\mathbb{A}_{\text{Uther}}$ by a reduct formation operation. In contrast to the depictions of the corresponding graphical configuration in Fig. 12.4, the PHON FIRST reduct of *Uther* in (13) contains the relation(s).

(13) The PHON FIRST reduct of $\mathbb{A}_{\text{Uther}}$:
$\beta_{\text{PF}} = \{\varepsilon\}$,
$\rho_{\text{PF}} = \{\langle\varepsilon,\varepsilon\rangle\}$,
$\lambda_{\text{PF}} = \{\langle\varepsilon, \textit{uther}\rangle\}$, and
$\xi_{\text{PF}} = \{\langle\texttt{component},\varepsilon,\varepsilon\rangle\}$.

The scientific purpose of relational abstract feature structures in linguistic theory is their use as conveniently structured mathematical entities which correspond to types of linguistic entities. The relational abstract feature structures admitted by a grammar are meant to constitute the predictions of the grammar (Pollard and Sag, 1994, p. 8).

In the context of our example, we are talking about one empirical prediction of the grammar $\langle\Sigma_2,\theta_2\rangle$, the prediction that the described language contains the utterance *Uther*. The exhaustive models mirror this prediction by containing (potentially multiple but isomorphic) *Uther* configurations. There is nothing else in the exhaustive models which has to do with this particular prediction of the grammar. The abstract feature structures admitted by the grammar predict six different types for this single expression. The six types are distinct, and they are unavoidable by construction if the grammar predicts the relational abstract feature structure which is an abstraction of a *Uther* configuration. The fundamental problem of the construction is that the well-formedness of $\mathbb{A}_{\text{Uther}}$ is only guaranteed by the well-formedness of all of its reducts. Hence we do not get a one-to-one correspondence between the types predicted by the grammar and the empirically observable expressions. Rather, it is the case that the abstract feature structures admitted by a grammar necessarily introduce a version of stranded structures, although there are no stranded monster structures among them as long as the grammar is a normal form grammar.[3]

I conclude that AFS fails to behave in the intended way. Even if one is willing to accept types of linguistic expressions as an appropriate target for linguistic theory, relational abstract feature structures are not adequate to make this approach to the theory of grammatical meaning technically precise.

Let us now turn to the second theory, MI. Pollard (1999) postulates that a formal grammar as a scientific theory should predict the grammatical utterance tokens of a natural language by specifying a set of structures which contains an idealized mathematical structure for each utterance token (and for nothing else). For two utterance tokens of the same expression there should only be one mathematical structure in the set. Moreover, the idealized mathematical structure should be structurally isomorphic to the utterance tokens it represents. This last condition is in fact much stronger than what (Pollard and Sag, 1994) asks from its linguistic types. Pollard and Sag's linguistic types merely stand in a relationship of conventional correspondence to utterance tokens. The conventional correspondence must be intuited by linguists without any further guidance with respect to the correctness of these intuitions from the meta-theory of linguistic meaning.

The most significant technical difference compared to AFS resides in how Pollard sets out to construct the mathematical idealizations of utterance tokens. Pollard's construction eschews relational abstract feature structures and consequently does not need the specialized feature structure satisfaction and admission relations of strictly feature structure based grammar formalisms. Instead, Pollard starts from the conventional grammar models of King (1999). From these standard models he proceeds to define *singly generated models* and then canonical representatives of singly generated models as mathematical idealizations of utterance tokens.

A singly generated model is a connected configuration under an entity which is actually a model of a grammar.

[3] Nothing substantial changes when we include the structure generated by the attribute EMBEDDED in the relational abstract feature structures. All four component sets of $\mathbb{A}_{\text{Uther}}$ as well as those of its five reducts become infinite, but the six feature structures remain distinct mathematical entities seemingly representing six different linguistic types.

In other words, a singly generated model has a topmost entity such that all other entities in the model are components of it. However, this is not yet the whole picture. Pollard defines the structures of interest as models together with their distinguished topmost entity. They are pairs, $\langle u, \langle \mathsf{U}_u, \mathsf{S}_u, \mathsf{A}_u, \mathsf{R}_u \rangle \rangle$, usually simply written as $\langle u, \mathsf{I}_u \rangle$.[4] The subscripts indicate that all entities in the universe U are components of u. We could say that I is a connected configuration under u which happens to be a model of a given grammar. Pollard then uses the distinguished entity in the configuration to define the canonical representative for each $\langle u, \mathsf{I}_u \rangle$ of the grammar. In essence, the entities in the canonical representatives are defined as equivalence classes of terms relative to the distinguished root entity. Not all details are relevant here,[5] the only important thing to note is that the standard model-theoretic technique of using terms of the logical language in the construction of a canonical model guarantees the uniqueness of each $\langle u, \langle \mathsf{U}_u, \mathsf{S}_u, \mathsf{A}_u, \mathsf{R}_u \rangle \rangle$ by the extensionality of the set-theoretic entities which serve as the elements of the universe U_u. As a result, Pollard manages to fix the canonical structure which stands for all isomorphically configured structures or utterance tokens. In order to have a name for them, I will henceforth call them *canonical representatives*. The collection of all canonical representatives of a grammar is the prediction of a grammar.

As in the investigation of AFS, I will focus on one prediction of $\langle \Sigma_2, \theta_2 \rangle$, the prediction that the utterance *Uther* will be judged grammatical. Although the structures of MI are defined quite differently from the set of relational abstract feature structures admitted by it, we will see immediately that AFS and MI share closely related problematic aspects.

Assume that we apply Pollard's method of constructing the canonical universes of Σ_2 interpretations as equivalence classes of Σ_2 terms. (14) shows schematically which canonical representatives Pollard's construction yields for the *Uther* configuration when it is applied to our exhaustive model. The subscripts indicate which entity of the exhaustive model of Fig. 12.3 is turned into the root entity of each of the six canonical representatives. By construction, each of the canonical representatives in (14) is a different set-theoretic entity. In brackets I mention the species of each root entity.

(14) a. $\langle u_{15}, \langle \mathsf{U}_{15}, \mathsf{S}_{15}, \mathsf{A}_{15}, \mathsf{R}_{15} \rangle \rangle$ (*u_word*)
b. $\langle u_{16}, \langle \mathsf{U}_{16}, \mathsf{S}_{16}, \mathsf{A}_{16}, \mathsf{R}_{16} \rangle \rangle$ (*nelist*)
c. $\langle u_{17}, \langle \mathsf{U}_{17}, \mathsf{S}_{17}, \mathsf{A}_{17}, \mathsf{R}_{17} \rangle \rangle$ (*elist*)
d. $\langle u_{18}, \langle \mathsf{U}_{18}, \mathsf{S}_{18}, \mathsf{A}_{18}, \mathsf{R}_{18} \rangle \rangle$ (*uther*)
e. $\langle u_{19}, \langle \mathsf{U}_{19}, \mathsf{S}_{19}, \mathsf{A}_{19}, \mathsf{R}_{19} \rangle \rangle$ (*cat*)
f. $\langle u_{20}, \langle \mathsf{U}_{20}, \mathsf{S}_{20}, \mathsf{A}_{20}, \mathsf{R}_{20} \rangle \rangle$ (*noun*)

It is immediately obvious that we observe here the same effect which we saw before with Pollard and Sag's utterance types. Since the *Uther* configuration contains six entities there are six distinct canonical representatives for it, although I assume that they would constitute one single prediction in Pollard's sense. The intended prediction seems to be that utterance tokens isomorphic to the *Uther* configuration are grammatical. In fact, for each n with $15 \leq n \leq 20$, all $\langle \mathsf{U}_n, \mathsf{S}_n, \mathsf{A}_n, \mathsf{R}_n \rangle$ in (14) are isomorphic, but this is not relevant in the construction. MI distinguishes between the corresponding entities in the universes because they are made of different equivalence classes of terms. Intuitively, the problem is that the entities are in different locations relative to their root entity, which entails that they are in a different equivalence class of terms defined on the root entity.[6]

I conclude that Pollard's construction fails to behave in the intended way. Pollard suggests that an HPSG grammar should be interpreted as specifying a set of canonical representatives such that no two members of the set are isomorphic, and utterance tokens of the language which are judged grammatical are isomorphic to one of the canonical representatives. Even if one is prepared to share Pollard's view of the goal of linguistics as a scientific theory, the particular construction proposed in (Pollard, 1999) is not suited to realize this conception without serious problems. For normal form grammars it introduces exactly the multiplicity of canonical representatives which it was designed to eliminate.

To sum up the preceding discussion, AFS and MI clearly fall short of the goals their proponents set for themselves. Neither Pollard and Sag's set of structures corresponding to linguistic utterance types nor Pollard's set of canonical representatives isomorphic to grammatical utterance tokens meets the intentions of their respective authors.

12.5 Minimal Exhaustive Models

I will now present an extension of King's theory of exhaustive models which avoids his problematic ontological commitment to possible utterance tokens, while retaining all other aspects of his model theory. At the same time, I also avoid the commitments to the ontological reality of utterance types or to the mathematical nature of the grammar models, which are characteristic of the metatheories (1) and (2). My starting point are the structural assumptions of normal form HPSG grammars, which I take to be independently motivated by the arguments in Section 12.3. For normal form grammars I define unique models which contain exactly one structure which is isomorphic to each utterance of a language considered well-formed by an ideal speaker of the language. This is, of course, what (1) and (2) essentially wanted to do, except

[4] The notation is explained in some detail in Section 12.5.

[5] They can be found in (Pollard, 1999, pp. 294–295) and even more explicitly in (Richter, 2004, pp. 208–210).

[6] It should be pointed out that the six interpretations in (14) are only isomorphic because we assume normal form grammars with an attribute EMBEDDED. However, without the EMBEDDED attribute we would run into the problems discussed in Section 12.3. In particular we would have stranded monster structures, and they would occur as canonical representatives which should correspond to possible utterance tokens, contrary to fact.

that I define *minimal exhaustive models* in such a way that I am not forced to make any commitments to the ontological nature of the structures in them. Given the philosophical intricacies of such commitments, I take this to be a highly desirable property of my proposal.

The goal is to characterize the meaning of grammars in terms of a set of structures, $\mathcal{M}$, which should have at least the following three properties: Each structure in $\mathcal{M}$ should have empirical consequences, i.e., there must be empirical facts which can falsify the predictions embodied by the structure; there should not be isomorphic copies of any empirically significant structure in the set of structures $\mathcal{M}$ assigned to each grammar; and finally, in accordance with one of Pollard's criteria, actual utterance tokens which are judged grammatical must be isomorphic to precisely one element in $\mathcal{M}$.

At first this small collection of desirable properties of $\mathcal{M}$ might seem arbitrary, even if every one of them can be individually justified. However, there is a way of integrating them with King's well-motivated theory of exhaustive models.

King's theory of grammatical truth conceives of language as a system of possible linguistic tokens. It claims that the system of possible tokens can be described as an exhaustive model of a grammar. The controversial aspect of this theory concerns the idea that language is a system of possible (i.e., actual and non-actual) tokens. Assume that we give up this aspect of King's theory. Instead we take an agnostic view toward language and say that we do not really know what it consists of. In our grammars we only make predictions about the discernible shapes of the empirical manifestations of language. We can operationalize this conception as follows: We want to write grammars such that whenever we encounter an actual utterance token, it will be judged grammatical if and only if there is an isomorphically structured connected configuration in an exhaustive model of the grammar. The connected configurations of interest will turn out to be the familiar connected configurations under unembedded signs. The choice of exhaustive model will not matter, since we are only concerned with the shape of the configurations, and we know that all shapes are present in any exhaustive model (by definition). However, since we are no longer after a system of possible tokens with an unknown number of isomorphic copies of configurations, we can be more precise about our choice of exhaustive model. It suffices to choose one which contains just one copy of each relevant connected configuration.

The theory of meaning we obtain from these considerations is a weakened form of King's theory. King says that a grammar is true of a natural language only if the language can be construed as a system of possible tokens, and the system of possible tokens forms an exhaustive model of the grammar. The theory proposed here as an alternative refrains from making such strong claims about the nature of language. It says that a grammar is true of a natural language only if each actual utterance token which is judged grammatical by an ideal speaker of the language is isomorphic to a maximal connected configuration in a minimal exhaustive model. The definitions of *maximal connected configurations* and *minimal exhaustive models* will be supplied directly below. Note that this condition endorses all arguments which King adduced to motivate exhaustive models, except for the ontological claim that the intended model is a system of possible (actual and non-actual) tokens.

Connected configurations in interpretations have been a leading intuitive concept since the first examples above. Their definition is straightforward. It presupposes the familiar RSRL signatures with a sort hierarchy $\langle \mathcal{G}, \sqsubseteq \rangle$, a distinguished set of maximally specific sorts $\mathcal{S}$, a set of attributes $\mathcal{A}$, an appropriateness function $\mathcal{F}$, and a set of relation symbols $\mathcal{R}$ whose arity is determined by a function $\mathcal{AR}$. Interpretations consist of a universe of objects U, a sort assignment function S which associates a symbol from $\mathcal{S}$ with each object in U, an attribute interpretation function A which treats each attribute symbol as the name of a partial function from U to U, and a relation interpretation function R which interprets each relation symbol as a set of tuples of the appropriate arity. Co^u_I is the set of those objects in U which can be reached from u by following a (possibly empty) sequence of attributes.

Definition 12.5.1. *For each signature* $\Sigma = \langle \mathcal{G}, \sqsubseteq, \mathcal{S}, \mathcal{A}, \mathcal{F}, \mathcal{R}, \mathcal{AR} \rangle$, *for each* Σ *interpretation* $\mathsf{I} = \langle \mathsf{U}, \mathsf{S}, \mathsf{A}, \mathsf{R} \rangle$,

$\langle \mathsf{U}', \mathsf{S}', \mathsf{A}', \mathsf{R}' \rangle$ *is a* connected configuration in I *iff*

1. $\mathsf{U}' \subseteq \mathsf{U}$,
2. *for some* $u' \in \mathsf{U}'$, $\mathsf{Co}^{u'}_\mathsf{I} = \mathsf{U}'$,
3. $\mathsf{S}' = \mathsf{S} \cap (\mathsf{U}' \times \mathcal{S})$,
4. $\mathsf{A}' = \mathsf{A} \cap (\mathcal{A} \times \{\mathsf{U}' \times \mathsf{U}'\})$,
5. $\mathsf{R}' = \mathsf{R} \cap \left(\mathcal{R} \times Pow\left(\bigcup_{n \in \mathbb{N}} \left(\overline{\mathsf{U}'}\right)^n \right) \right)$.

Certain connected configurations in interpretations are of special interest to us. These are connected configurations which are not properly contained within other connected configurations in their interpretation. I will call them *maximal*:

Definition 12.5.2. *For each signature* Σ, *for each* Σ *interpretation* $\mathsf{I} = \langle \mathsf{U}, \mathsf{S}, \mathsf{A}, \mathsf{R} \rangle$,

$\langle \mathsf{U}', \mathsf{S}', \mathsf{A}', \mathsf{R}' \rangle$ *is a* maximal connected configuration in I *iff*

$\langle \mathsf{U}', \mathsf{S}', \mathsf{A}', \mathsf{R}' \rangle$ *is a connected configuration in* I,
and for some $u' \in \mathsf{U}'$:
$\mathsf{Co}^{u'}_\mathsf{I} = \mathsf{U}'$, *and for every* $u'' \in \mathsf{U}$, $\mathsf{Co}^{u'}_\mathsf{I} \not\subset \mathsf{Co}^{u''}_\mathsf{I}$.

There are three maximal connected configurations in the interpretation of Fig. 12.1. Their topmost elements are the *phrase* entity 16, which is the topmost entity in the connected configuration with the phonology *Uther walks*; the *word* entity 30, which is the topmost entity in the connected configuration with the phonology *Uther*; and the *word* entity 19, which is the topmost entity in the connected configuration with the phonology *walks*.

We can prove important properties of maximal connected configurations in models of normal form grammars: No two of them overlap. Each of them contains exactly one *u_sign* entity, which guarantees that they are empirical structures. Each entity in a model actually belongs to a maximal connected configuration, which ensures the empiricity of all entities. Every *u_sign* entity is contained in a maximal connected configuration, which guarantees that maximal connected configurations indeed capture all empirically relevant predictions without missing any. From now on I refer to maximal connected configurations in models of normal form grammars as *u-sign configurations*. The u-sign configurations in models of our grammars constitute the empirical predictions of the grammars.

I define *minimal exhaustive grammar models* as exhaustive models which contain exactly one copy of each possible u-sign configuration.

Definition 12.5.3. *For each signature* Σ, *for each* Σ-*theory* θ, *for each exhaustive* $\langle\Sigma,\theta\rangle$ *model* I,

I *is a* minimal exhaustive $\langle\Sigma,\theta\rangle$ model *iff*

for each maximal connected configuration I_1 *in* I, *for each maximal connected configuration* I_2 *in* I*:*

if I_1 *and* I_2 *are isomorphic then* $\mathsf{I}_1 = \mathsf{I}_2$.

The exhaustive $\langle\Sigma_2,\theta_2\rangle$ model of Fig. 12.3 is an example of a minimal exhaustive grammar model. It contains exactly one copy of each u-sign configuration predicted by the grammar $\langle\Sigma_2,\theta_2\rangle$.

According to the properties of u-sign configurations, a minimal exhaustive model of a normal form grammar is partitioned into separate u-sign configurations. Each pair of u-sign configurations in it is structurally distinct and thus constitutes a different prediction of the grammar. Since all connected configurations in these models are u-sign configurations, they do not contain anything which is empirically vacuous.

With my construction I have not made any ontological commitments. I have claimed that the internal structure of actual utterance tokens can be discovered, and that this structure is mirrored precisely in u-sign configurations in minimal exhaustive grammar models. This did not presuppose saying anything about the ontology of linguistic objects. It was not even necessary to say what kinds of entities populate the minimal exhaustive models.

12.6 Concluding Remarks

Should there be any concern about the undetermined nature of the entities in minimal exhaustive models, or a preference for mathematical models, it is possible to pick out one mathematical model and fix it as the intended minimal exhaustive model of a given normal form grammar. The architecture of minimal exhaustive models of normal form grammars suggests strongly how to do this. Since the minimal exhaustive models are populated by a collection of u-sign configurations, and since the unique *u_sign* entity in each u-sign configuration contains all other elements of the configuration as its components, it is quite natural to define the entities in the u-sign configurations as equivalence classes of paths which lead to them from their individual *u_sign*. This of course is essentially Pollard's construction of canonical representatives, except that I avoid the multiplicity of representatives for one and the same prediction because my mathematical idealizations do not consist of pairs of entities and configurations. Instead, I exploit the special properties of the models of normal form grammars and am thus able to make do with bare u-sign configurations.

But although the construction of minimal exhaustive models from mathematical entities is simple, I am not aware of any convincing argument for them. In my opinion, DEFINITION 12.5.3 completes the explanation of the meaning of normal form HPSG grammars.

Bibliography

King, Paul J. (1999). Towards Truth in Head-driven Phrase Structure Grammar. In Valia Kordoni, ed., *Tübingen Studies in Head-Driven Phrase Structure Grammar*, Arbeitspapiere des SFB 340, Nr. 132, Volume 2, pp. 301–352. Eberhard Karls Universität Tübingen.

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

Pollard, Carl J. (1999). Strong generative capacity in HPSG. In Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol, eds., *Lexical and Constructional Aspects of Linguistic Explanation*, pp. 281–297. CSLI Publications.

Richter, Frank (2004). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Phil. dissertation (2000), Eberhard Karls Universität Tübingen.