Encoding Information Packaging Information in HPSG *

Suresh Manandhar
Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland
Suresh.Manandhar@edinburgh.ac.uk

May 19, 1998

1 An alternative encoding in HPSG

In this commentary, I provide what I believe to be a simpler and possibly more elegant alternative to the HPSG based encoding presented in section section 3.2 of [EV94]. The principal difference is firstly to treat *focus, ground, link* and *tail* as values of the INFO-STRUCT feature as opposed to being feature symbols themselves. Secondly focus inheritance discussed in section 3.2.2 of [EV94] is treated by employing Reape's *domain-union* mechanism [Rea94] thereby providing a mechanism for stating further constraints between focus and word-order. This would also provide the starting point for the exploration of the interaction between focus and word-order which is deemed as future work in [EV94]. Of course the specification of such principles is beyond the scope of this very short commentary.

We begin by assuming that *focus, ground, link* and *tail* are ordered types ordered according to the hierarchy depicted in figure 1. Similarly, for *accent* marking on the PHON attribute we assume the flat hierarchy in figure 1.

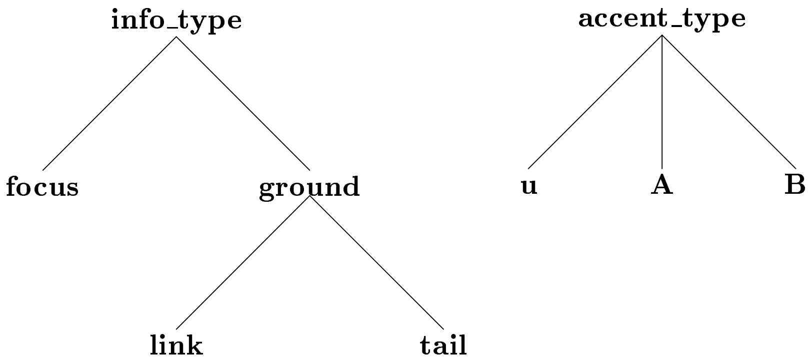


Figure 1: Focus and accent hierachy

We assume that the attribute INFO-STRUCT is defined for every *sign*. We shall be vague about whether INFO-STRUCT is part of the SYNTAX, SEMANTICS or CONTEXT attribute. For our purposes it suffices to assume that INFO-STRUCT is a top-level attribute appropriate for every *sign*.

*I am thankful to both Elisabet Engdahl and Enric Vallduví for helping me undertand their paper [EV94] upon which this commentary is based.

Following, Reape [Rea94] we assume that the DOM feature (short for DOMAIN) is appropriate as the value of the SYNSEM|LOC path for every sign. The DOM feature is employed by Reape to build word-ordering domains locally during a derivation step. For our purposes, it suffices to assume the following *Domain Principle* specified for head-complement structures:

(1) *Domain Principle*:
The value of the SYNSEM|LOC|DOM feature of every sign is the **domain union** of the values of SYNSEM|LOC|DOM feature of the HEAD-DTR and that of every sign in the COMP-DTRS list.

We refer the reader to [Rea94] for the definition of *domain union* operation which concatenates two domains to build a larger one.

Our Domain Principle is a simplification of Reape's Domain Principle which permits arguments to declare (by the UNION $\pm$ feature) whether their domains can be inherited or not. This simplification is purely for a simpler presentation.

With the Domain Principle in place the INFO_STRUCT values of head daughter and complement daughters are automatically inherited by the mother.

Now we need a principle to enforce focus constraints at a phrasal level. For this we assume that the INFO-STRUCT value of the mother subsumes the INFO-STRUCT values of every sign in its domain. This is stated as the Focus Inheritance Principle given below.

(2) *Focus Inheritance Principle*:
The value of the INFO-STRUCT attribute of a sign **subsumes** (*i.e.* is at least as general as) the value of the INFO-STRUCT attribute of every sign in the SYNSEM|LOC|DOM list[1].

Once consequence of the Focus Inheritance Principle is that if a constituent is marked as *ground* then every subconstituent (*i.e.* every element of the DOM list) too is at least as specific as *ground*. This would allow for the possibility that some subconstituents may be more specifically marked as *link* or *tail*.

The Focus Inheritance Principle is intended to be a language universal principle and as we shall see additional language specific principles will be needed.

One practical problem with our formulation of the Focus Inheritance Principle is the use of subsumption constraints. It is known that subsumption constraints cause *undecidability* [DR90] and is generally not available in current typed feature formalisms. But the Focus Inheritance Principle can be approximated by the following set of conditional constraints.

(3) *Focus Inheritance Principle*:

$$sign\begin{bmatrix}\text{INFO-STRUCT} & focus\\ \text{SYNSEM}|\text{LOC}|\text{DOM} & \boxed{1}\end{bmatrix} \Rightarrow all\_info\_struct(\boxed{1}, focus).$$

$$sign\begin{bmatrix}\text{INFO-STRUCT} & link\\ \text{SYNSEM}|\text{LOC}|\text{DOM} & \boxed{1}\end{bmatrix} \Rightarrow all\_info\_struct(\boxed{1}, link).$$

$$sign\begin{bmatrix}\text{INFO-STRUCT} & tail\\ \text{SYNSEM}|\text{LOC}|\text{DOM} & \boxed{1}\end{bmatrix} \Rightarrow all\_info\_struct(\boxed{1}, tail).$$

$$sign\begin{bmatrix}\text{INFO-STRUCT} & ground\\ \text{SYNSEM}|\text{LOC}|\text{DOM} & \boxed{1}\end{bmatrix} \Rightarrow ground\_all\_info\_struct(\boxed{1}).$$

$$all\_info\_struct([], \boxed{2}).$$
$$all\_info\_struct([\begin{bmatrix}\text{INFO-STRUCT } \boxed{2}\end{bmatrix} \mid \boxed{1}], \boxed{2}) :-$$
$$\quad all\_info\_struct(\boxed{1}, \boxed{2}).$$
$$ground\_all\_info\_struct([]).$$
$$ground\_info\_struct([\begin{bmatrix}\text{INFO-STRUCT } ground\end{bmatrix} \mid \boxed{1}]) :-$$
$$\quad ground\_info\_struct(\boxed{1}).$$

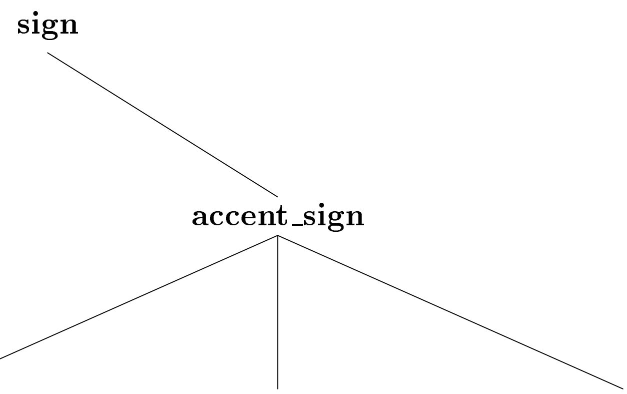
[1] An alternative to the use of the subsumption relation is to require that the INFO-STRUCT value of the mother is the **generalisation** of the INFO-STRUCT values of every sign in the SYNSEM|LOC|DOM list. We do not explore this possibility in this report.

Stated in this manner, the Focus Inheritance Principle states token identity between the INFO-STRUCT values of the mother and all the signs in the DOMAIN list just in case the INFO-STRUCT value of the mother is either *focus*, *link* or *tail*. On the other hand, it enforces type identity between the INFO-STRUCT values of the mother and all the signs in the DOMAIN list just in case the INFO-STRUCT value of the mother is *ground*. The last condition permits further specialisation of the *ground* type to either *link* or *tail*.

2 Prosodic marking

The co-restrictions between accent information and focus information given in section 3.2.1. of [EV94] can then be stated in this modified framework by assuming that every *sign* further specialises into *accent_sign* where the definition of *accent_sign* is given below.

(4)



$$u\_type \begin{bmatrix} \text{PHON|ACCENT} & u \\ \text{INFO-STRUCT} & info\_type \end{bmatrix} \quad a\_type \begin{bmatrix} \text{PHON|ACCENT} & A \\ \text{INFO-STRUCT} & focus \end{bmatrix} \quad b\_type \begin{bmatrix} \text{PHON|ACCENT} & B \\ \text{INFO-STRUCT} & link \end{bmatrix}$$

We further assume a feature appropriateness mechanism which automatically enforces the type *u_type* if the value of the ACCENT feature is u and similarly for types *a_type* and *b_type*.

3 Language specific principles

Here we outline a reformulation of the INFO-STRUCT inheritance principles for English given in [EV94]. We state these as two independent principles. The first principle states that for unmarked accent the value of the INFO-STRUCT feature can be either *focus* or *tail* (thus ruling out the possibility for *link*). The second principle states that if the most *oblique* daughter has unmarked accent then the value of its INFO-STRUCT feature must be *tail*. Thus, our second condition can be seen as further constraining the first condition.

The first principle can be formulated most easily by specialising the definition of *u_type* to the one given below.

(5) *English Focus Principle A*

$$u\_type \begin{bmatrix} \text{PHON|ACCENT} & u \\ \text{INFO-STRUCT} & focus \sqcup tail \end{bmatrix}$$

The second principle can be formulated in an informal manner as given below.

(6) *English Focus Principle B*

$$word \left[\text{SYNSEM|LOC|CAT|SUBCAT} \left\langle \ldots, \left[\text{PHON|ACCENT}\ u\right] \Rightarrow \left[\text{INFO-STRUCT}\ tail\right] \right\rangle \right]$$

The description $\left[\text{PHON|ACCENT}\ u\right] \Rightarrow \left[\text{INFO-STRUCT}\ tail\right]$ is a conditional description which states that if the value of the ACCENT feature is of type a then the value of its INFO-STRUCT feature must be of

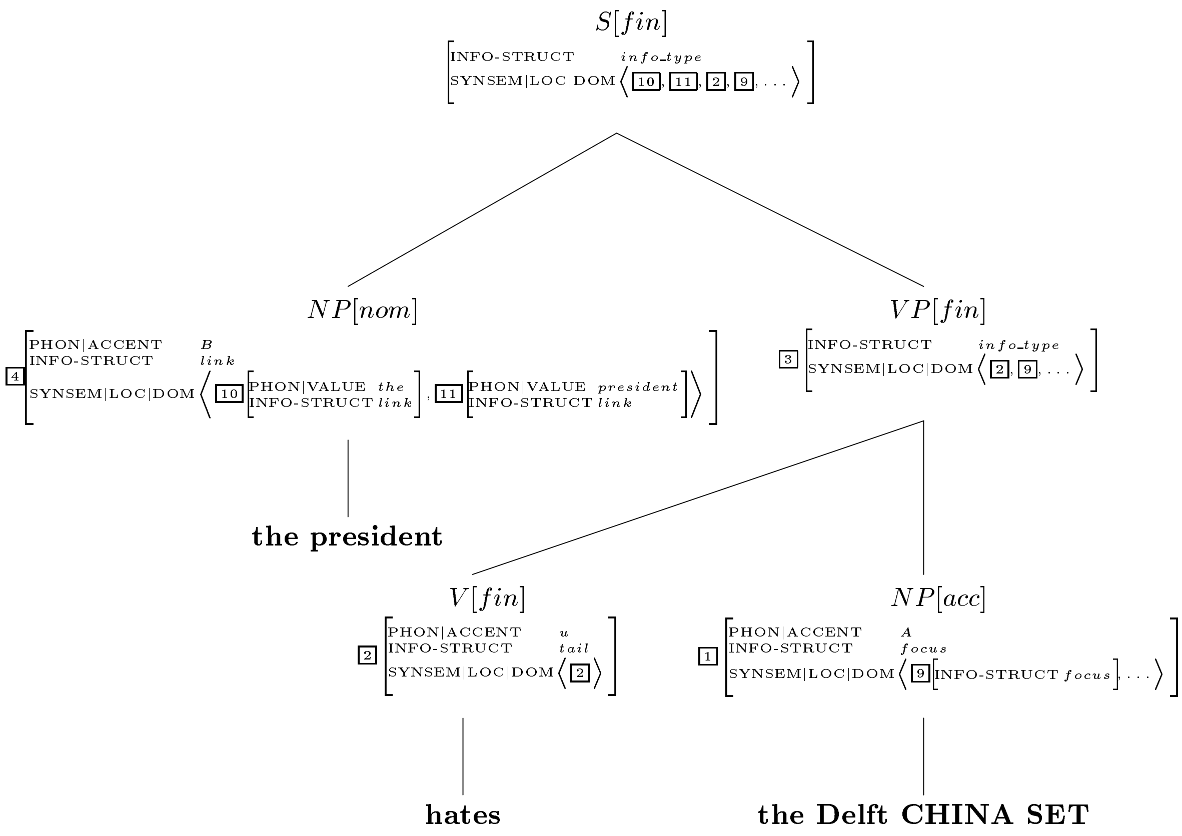


Figure 2: Alternative Derivation

type *tail*. A more formal version can be easily formulated for instance by employing a relational dependency (to access the last element of a list).

With the language universal Focus Inheritance Principle and the language specific Principles in place, the example derivation of an *object NP focus* construction given in example (42) in section 3.2.2 of [EV94] would look like the derivation given in figure 2.

Some explanation is needed for figure 2. The structure $\left\langle \boxed{9}\left[\text{INFO-STRUCT } \textit{focus}\right], \dots \right\rangle$ corresponds to the derivation of the NP *the Delft CHINA SET*. The whole NP is marked for accent A and hence is focal since this sign is of type *a_type*[2]. The sign corresponding to the verb *hates* has unmarked accent and hence by English Principle A can be either a *tail* or a *focus*. We choose it to be *tail*. Note that in theory this choice could be made by some pragmatic component of the grammar.

Next our Domain Union Principle requires that *VP*[*fin*] (indicated by index $\boxed{3}$ in figure 2) inherit the domains of its head daughter (in this case *V*[*fin*]) and the complement daughter (in this case *NP*[*acc*]). However, our Focus Inheritance Principle forces the INFO-STRUCT value of *VP*[*fin*] to be *info_type* since this is the only common super-type of both *tail* and *focus* (see figure 1). This contrasts with the approach taken in [EV94] where *VP*[*fin*] would contain a FOCUS feature containing *NP*[*acc*] and a GROUND|TAIL feature containing *V*[*fin*]. However, note that within the current approach every *focus* sign (resp. *link* and *tail*) in the DOM list can be extracted into a separate list (for instance by employing a relational dependency mechanism). For the moment,

[2]We have little to say about the phrasal inheritance of the ACCENT feature. We assume that an appropriate mechanism is in place and the interested reader should check the references cited in [EV94]

there does not appear to be a need for this separate representation.

Finally, the DOM feature of *S[fin]* contains the domains of *NP[nom]* and *VP[fin]*. Again, the values in the DOM list of *NP[nom]* corresponds to the NP *the president*. *NP[nom]* being marked for accent B can only be a *link*. This means that by our Focus Inheritance Principle, *S[fin]* can only be of type *info_type*.

It can be seen that other example sentences reproduced below from the ones given in section 3.2.2 of [EV94] too are correctly predicted.

(7) a. The president [$_F$ HATES] the Delft china set.
b. The president [$_F$ hates the Delft CHINA SET].
c. *The president [$_F$ HATES the Delft china set].

Example (7)a is a link-focus-tail construction and is licenced by both English Focus Principle A and Principle B. Example (7)b is licenced since the verb has accent u and is allowed to be focal (by Principle A). The Focus Inheritance Principle then licences the whole VP to be the focus. Example (7)c is ruled out since Principle B rules out the object NP to be focal assuming that it has unmarked accent. The Focus Inheritance Principle then blocks the whole VP being the focus.

Similarly thetic sentences where the subject is associated with accent A given in section 3.2.2. of [EV94] too are correctly predicted:

(8) a. [$_F$ The PRESIDENT] called.
b. [$_F$ The PRESIDENT called].
c. *[$_F$ The president CALLED].

Focus inheritance is permitted in sentence (8)b but is blocked in sentence (8)c since the subject NP also functions as the most oblique complement making Principle B applicable.

4 Conclusions

The HPSG encoding presented in this commentary can be thought of as a generalisation and simplification of the encoding given in section 3.2 of [EV94]. Furthermore, our encoding provides a general-purpose representation for stating further conditions between word-order and focus.

The approach advocated in this short commentary makes it possible to begin implementation of a test grammar within current typed feature systems such as CUF [DD93] or ALE [Car93]. This should highlight potential problems with the current approach and thereby suggest improvements.

References

[Car93] Bob Carpenter. ALE:Attribute Logic Engine Users Guide, Version β. Technical report, Carnegie Mellon University, Pittsburgh, PA 15213, 1993.

[DD93] Jochen Dörre and Michael Dorna. CUF: A Formalism for Linguistic Knowledge Representation. Dyana-2 deliverable, IMS, Stuttgart, Germany, August 1993.

[DR90] Jochen Dörre and William C. Rounds. On subsumption and semiunification in feature algebras. In *5th Annual IEEE Symposium on Logic In Computer Science*, pages 300–310, Philadelphia, PA, June 1990. IEEE Computer Society Press.

[EV94] Elisabet Engdahl and Enric Vallduví. Information Packaging and Grammar Architecture: A constraint-based approach. In Elisabet Engdahl, editor, *Integrating information structure into constraint-based and categorial approaches*. DYANA-2, Deliverable R1.3.B, Amsterdam: Institute for Language, Logic, and Computation, September 1994.

[Rea94] Mike Reape. Getting Things in Order. In Wietske Sijtsma and Arthur van Horck, editors, *Discontinuous Constituency*. Berlin: Mouton de Gruyte, 1994. To appear.