Representing information structure in a formal grammar of Danish[*]

Patrizia Paggio

Center for Sprogteknologi
University of Copenhagen
`patrizia@cst.dk`

Abstract. This paper presents a proposal for the integration of information structure in a unification-based grammar of Danish. Three information structure features – topic, focus and background – are defined, and it is shown how they are instantiated in a number of different grammar constructions. Prosodic, syntactic and information structure constraints characterising the various constructions are represented as typed features following Head-driven Phrase Structure Grammar (HPSG), and the constructions themselves are ordered in a type hierarchy. The proposed approach modifies and extends earlier HPSG-based accounts by i. defining a new interface between information structure and semantics, and ii. integrating information structure as a dimension of phrasal and clausal grammar constructions.

1 Introduction

In previous works [1] [2] I have described some of the peculiarities of Danish with respect to the way in which focality or topicality are expressed. The most notable features of this language in this respect are the lack of sentence accent, and the role played by word order and by syntactically marked constructions such as clefting and topicalisation. In this paper I will not focus much on these aspects, and instead address the issue of how information structure can be represented in a formal, unification-based grammar of the language within the framework of HPSG [3]. I propose a model in which information structure is formalised as a dimension in a hierarchy of constructions where it interacts with prosodic, syntactic and semantic properties of phrasal and clausal types.

The model is based on three fundamental features. The first is *focus*, which corresponds to the sentence elements that indicate a modification of the pragmatic information built by the overall discourse. The part of the sentence which is not in focus is the *background*, which can be considered peripheral to the construction of the discourse, and serves the purpose of specifying or restricting the way in which the focal information must contribute to it. Finally, the third feature is the *topic*, which can be characterised as the entity which the sentence expresses pertinent information about, thereby increasing the hearer's knowledge of it. As we will see, not all three categories are necessarily present in every sentence, although the focus always is. Depending on which of the category is instantiated, different information structure types can be defined, and their relation to grammar constructions formalised as structure sharing patterns between focus, topic and background on the one hand, and features pertaining to the other dimensions of the sign on the other.

In Sect. (2) the feature structure corresponding to the information structure of a sentence is introduced, and its relation to the semantics is discussed. I argue here that my approach extends earlier HPSG-based accounts by defining a new interface between information structure and semantics. In Sect. (3) a formal account of the four basic information structure types is given, followed by a description of local features in Sect. (4) and an explanation of the principles that govern the instantiation of information structure features in a syntactic derivation in Sect. (5). In Sect. (6) an account of the interaction between information structure types and grammar constructions is provided, and finally Sect. (7) contains the conclusions.

---

2 Representing information structure

In HPSG the basic unit of linguistic representation is the sign, which is represented as a complex feature structure conflating all its phonological, syntactic, semantic and pragmatic properties.

(1) $\begin{bmatrix} sign \\ \text{PHONOLOGY } list(phonstring) \\ \text{SYNSEM } synsem \\ \text{QSTORE } set(quantifiers) \\ \text{RETRIEVED } list(quantifier) \end{bmatrix}$

This representational style makes HPSG, and more in general grammar formalisms based on typed feature structures, attractive frameworks in which to capture the interaction between information structure and other aspects of the grammar. In such a framework, the contribution made by the various words to the information structure of a sentence is represented by means of features belonging to the pragmatic description of the sign, while the necessary links between these pragmatic features and attributes pertaining to the other linguistic dimensions are ensured by structure sharing.

To represent information structure, I will introduce the attribute INFOSTR as part of the context (CTXT) of the linguistic sign. The value of INFOSTR is a feature structure with the three attributes TOPIC, FOCUS and BG. Each of them takes as value a list of semantic indices[1].

(2) $\begin{bmatrix} \text{INFOSTR} & \begin{bmatrix} \text{TOPIC (list of indices)} \\ \text{FOCUS (list of indices)} \\ \text{BG (list of indices)} \end{bmatrix} \end{bmatrix}$

By means of the semantic indices we can let the information structure features refer to the elements that make up the semantic representation of the sentence, for which I rely on Minimal Recursion Semantics (MRS) [4] [5]. In MRS, the two main components of the semantics of a sentence are i. an unordered bag of relations, where each relation can be referred to by a label and can in turn refer to a number of argument variables (or indices), and ii. a number of scope assignments over pairs of relations. One of the strengths of MRS is the fact that the semantic representations may be underspecified in case of scope ambiguity, for example when several quantifiers are present. I will not, however, go into detail with this property here. In addition to the relation bag and the scope assignments, an MRS representation also contains an LTOP handle referring to the relation with the highest scope, and an instance or event INDEX introduced by the semantic head, analogous to lambda. For example, let us consider the following Danish sentence:

(3) Manden så barnet.
    Man-the saw child-the[2].
    'The man saw the child.'

An MRS representation of this sentence will be:

```
[LTOP:h1
 INDEX: e2
 RELS: <
   ["_se_v_rel"
    LBL:h8
    ARG0:e2
    ARG1:x4
    ARG2:x3],
```

---

[1] Set valued features would in fact be more appropriate, since the mutual order of the indices is irrelevant. However, sice sets are quite problematic to implement, lists have been chosen instead.

[2] In Danish the definite article is affixed to the noun in simple NPs.

```
["_barn_n_rel"
 LBL:h5
 ARG0:x3]
["_mand_n_rel"
 LBL:h6
 ARG0:x4]
[message_m_rel
 LBL:h1
 MARG:h7]>
  HCONS: <h7 qeq h8> ]
```

In this representation, the label (LBL) value of the *message* relation at the bottom of the relation list (RELS) is also the top handle (LTOP) of the sentence, whilst its argument (MARG) is equalled to the label of the relation introduced by the main verb via the *qeq* ('equality modulo quantifiers') constraint in the HCONS list. Furthermore, the verbal relation has three arguments referring to an event and to the instances associated with the relations *barn* (child) and *mand* (man), respectively.

The most obvious reading of this sentence in information structure terms is as a *topic-comment* construction, where the VP constitutes the comment to the topical subject. An alternative term well-known from the generative tradition is *wide focus*:

(4) $[_T$ Manden] $[_F$ så barnet].

On this reading, I suggest that the information structure part of the sign should look as follows:

(5)
$$\left[\text{INFOSTR} \begin{bmatrix} \text{TOPIC} \langle \text{x4} \rangle \\ \text{FOCUS} \langle \text{x3, e2} \rangle \\ \text{BG} \langle \rangle \end{bmatrix} \right]$$

The values of the information structure features coincide with indices in the MRS representation. The TOPIC coincides with the index associated with the subject, whilst the FOCUS is a list of two indices corresponding to the event introduced by the main verb and the instance corresponding to the object. The BG is an empty list. I believe this is a simple and elegant way of expressing the relation between information structure and content which I haven't seen attempted before. Its novelty lies not only in its use of MRS, but also in the fact that it solves a fundamental problem associated with the interpretation of wide focus.

To discuss this issue, let us look at how information structure has been represented in earlier works within the HPSG framework. In [6], information structure features are structure-shared with feature sorts containing syntactic, phonological and semantic information. For instance, the FOCUS feature of an accented sign is structure-shared with the sign itself, thereby producing a cyclic structure as shown below:

(6)
$$\boxed{1}\begin{bmatrix} \text{PHON} \,|\, \text{ACCENT A} \\ \text{INFO-STR} \,|\, \text{FOCUS} \,\boxed{1} \end{bmatrix}$$
(Engdahl and Vallduví 1996:11)

The same approach is also followed more recently by [7]. However, Engdahl and Vallduví argue themselves for a modification of their proposal where information structure features are structure-shared with content features rather than the whole sygn. In [8], Engdahl has in fact adopted this new approach, and the feature structure for an accented English word has become:

(7)
$$\begin{bmatrix} word \\ \text{PHON} \,|\, \text{ACCENT A} \\ \text{CONTENT} \,\boxed{1} \\ \text{INFO-STRUCT} \,|\, \text{FOCUS} \,\boxed{1} \end{bmatrix}$$

In Fig. (1), the representation proposed by the author for the wide VP focus interpretation of the sentence "John plays rugby" is shown in a slightly simplified version. Note that this English sentence displays the same information structure as the Danish example in (4), except for the fact that the subject is labelled LINK instead of TOPIC.
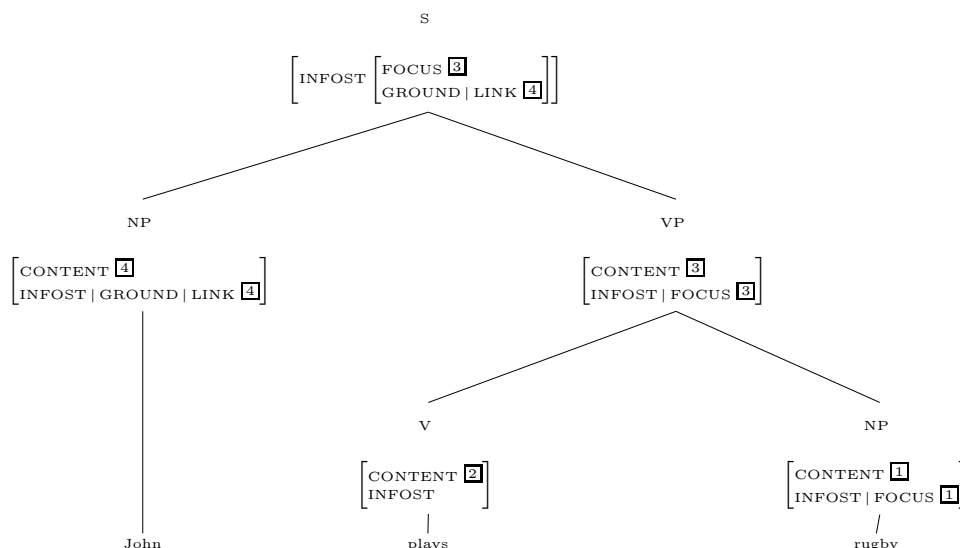
S

$$\left[\text{INFOST}\ \begin{bmatrix}\text{FOCUS}\ \boxed{3}\\ \text{GROUND} \mid \text{LINK}\ \boxed{4}\end{bmatrix}\right]$$

NP

$$\begin{bmatrix}\text{CONTENT}\ \boxed{4}\\ \text{INFOST} \mid \text{GROUND} \mid \text{LINK}\ \boxed{4}\end{bmatrix}$$

VP

$$\begin{bmatrix}\text{CONTENT}\ \boxed{3}\\ \text{INFOST} \mid \text{FOCUS}\ \boxed{3}\end{bmatrix}$$

V

$$\begin{bmatrix}\text{CONTENT}\ \boxed{2}\\ \text{INFOST}\end{bmatrix}$$

NP

$$\begin{bmatrix}\text{CONTENT}\ \boxed{1}\\ \text{INFOST} \mid \text{FOCUS}\ \boxed{1}\end{bmatrix}$$

John   plays   rugby

Fig. 1. Wide VP focus. Adapted from (Engdahl 1999:13)

The representation proposed in the figure is similar to mine in the sense that information structure features are structure-shared with content features. However, whereas in my approach the contents of verb and complement are simply *collected* in a list of indices, in Engdahl's proposal the focus is directly identified with the content of the VP node, and the verb does not make any contribution to information structure: its information structure features are left unspecified. The question now is, what does the content of the VP consist of? Is it different from just a list of the contents of the daughters of the VP node?

Engdahl's proposal is in fact problematic because in HPSG the content of the VP cannot be distinguished from the content of the whole sentence: the HPSG semantics principle set forth in [3] states in fact that in head-complement phrases, the content of a mother is token-identical with the content of the daughter. This principle is adhered to in versions of HPSG making use of MRS [9] as I do here. The consequence this strategy has for the treatment of wide focus, is that if the focus is structure-shared with VP content, it will always include the subject of the sentence. In other words, VP focus cannot be distinguished from S focus.

The problem is noted by Cooper [10], who in fact uses it as an argument for proposing a more elaborate linguistic representation (based on *record types* rather than feature types) that includes functions in its formal apparatus, thus building on the Montagovian tradition (e.g. [11]), where the content of a VP is a function of the content of the arguments. This is certainly an interesting avenue to pursue. In this study, however, I will stay with the basic MRS language, which can be considered a kind of standard in the HPSG community. Therefore, VP focus will be represented by letting the focus domain span over a collection of semantic indices corresponding to the verb and its complements. In [2], I have argued that sentence accent cannot be used in Danish to find the focus, and that non-pronominal complements and adjuncts in the focus domain must generally be accented. Also on this ground, then, it seems reasonable to represent focus as a collection of semantic indices corresponding to independently accented constituents.

In Fig. (2) the syntactic tree corresponding to the analysis of example (4) is shown with an indication of the information structure values. The variables $\boxed{1}$, $\boxed{2}$, $\boxed{3}$ show structure sharing

between the features in the tree. The ultimate values of these variables will correspond to instances or events as shown in (5).
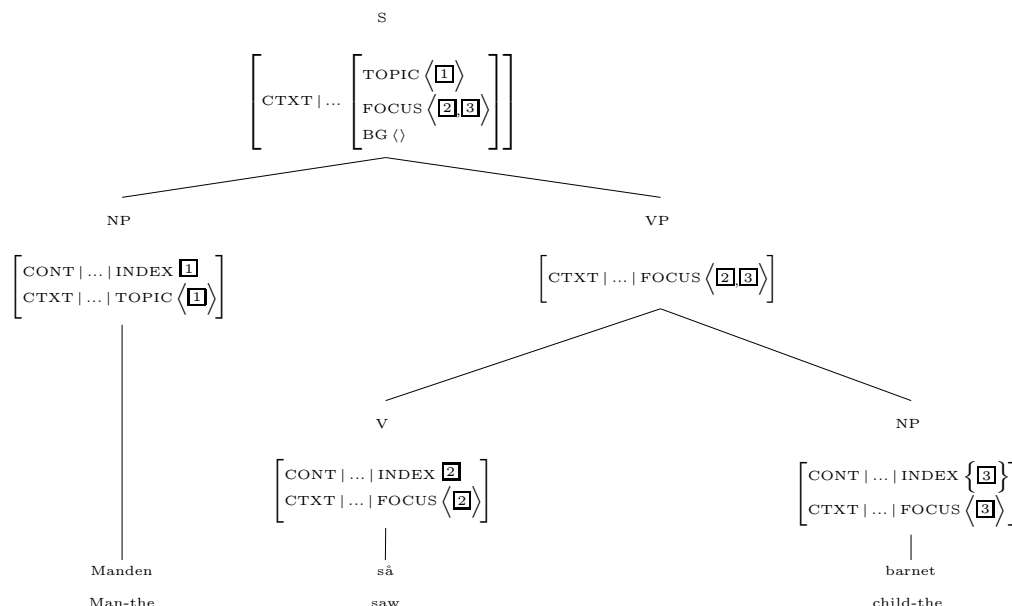

Fig. 2. Collecting indices in wide VP focus

Having established the relation between the information structure features and the content expressed in the MRS representations, let us now discuss the mechanisms that account for this relationship in the grammar. In summary, four different kinds of constructs are relevant and will be described below. They are:

- information structure types;
- topic, focus and background local features;
- inheritance principles;
- construction types.

3 Information structure types

Given the fact that a sentence always has a focus, whilst topic and background may or may not be present, we can posit four different types of information structure, which can be instantiated in grammar constructions in different ways. Below are examples in which the same event – the fact that a known group of children have eaten icecream – is reported by means of the four possible information structures:

(8) a. (Hvad har børnene lavet?) [$_T$ De] [$_F$ har spist is].
       'What have the children done? They have eaten icecream.'
   b. (Hvad har børnene spist?) [$_T$ De] [$_{BG}$ har spist] [$_F$ is].
       'What have the children eaten? They have eaten icecream.'
   c. (Hvem har spist is?) [$_F$ Børnene] [$_{BG}$ har spist is].
       'Who has eaten icecream? The children have eaten icecream.'
   d. (Var det sjovt?) [$_F$ Ja, børnene har spist is].
       'Was it fun? Yes, the children have eaten icecream.'

Working with typed feature structures gives us the possibility of structuring linguistic knowledge into type hierarchies. This applies also to the information structure types: they can be organised in a hierarchy defining the constraints that apply to each type as shown in Fig. (3).
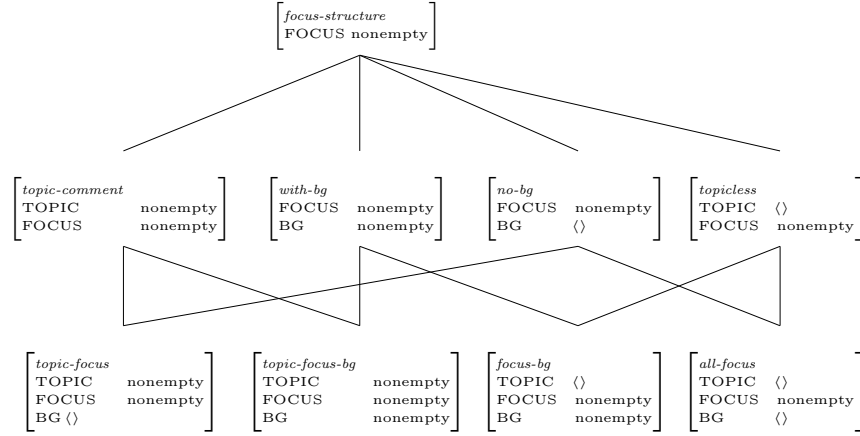
$$
\begin{bmatrix} \textit{focus-structure} \\ \text{FOCUS nonempty} \end{bmatrix}
$$

$$
\begin{bmatrix} \textit{topic-comment} \\ \text{TOPIC} \quad \text{nonempty} \\ \text{FOCUS} \quad \text{nonempty} \end{bmatrix}
\quad
\begin{bmatrix} \textit{with-bg} \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG} \quad \text{nonempty} \end{bmatrix}
\quad
\begin{bmatrix} \textit{no-bg} \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG} \quad \langle \rangle \end{bmatrix}
\quad
\begin{bmatrix} \textit{topicless} \\ \text{TOPIC} \quad \langle \rangle \\ \text{FOCUS} \quad \text{nonempty} \end{bmatrix}
$$

$$
\begin{bmatrix} \textit{topic-focus} \\ \text{TOPIC} \quad \text{nonempty} \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG } \langle \rangle \end{bmatrix}
\quad
\begin{bmatrix} \textit{topic-focus-bg} \\ \text{TOPIC} \quad \text{nonempty} \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG} \quad \text{nonempty} \end{bmatrix}
\quad
\begin{bmatrix} \textit{focus-bg} \\ \text{TOPIC} \quad \langle \rangle \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG} \quad \text{nonempty} \end{bmatrix}
\quad
\begin{bmatrix} \textit{all-focus} \\ \text{TOPIC} \quad \langle \rangle \\ \text{FOCUS} \quad \text{nonempty} \\ \text{BG} \quad \langle \rangle \end{bmatrix}
$$

Fig. 3. Constraints on information structure types

Of course, the relation between information structure types and the grammar constructions in which they come into play is not a one-to-one mapping. For example, a topic-focus information structure can be instantiated in Danish by an ordinary declarative clause or a topicalised sentence:

(9) a. [$_T$ Han] [$_F$ så barnet].
        'He saw the child.'
     b. [$_T$ Ham] [$_F$ vil barnet ikke tale med].
        Him will child-the not speak with.
        'The child will not speak with him.'

In fact, the realisation of information structure types in sentence constructions is the result of the application of both local constraints and more general information structure principles which will be described in the next sections.

4 Topic, focus and background features

In order to be able to express constraints on the information structure value of words and phrases, let us start by defining local features relevant for topic, focus and background signs. For instance in the local features for focus, the index value is structure-shared with an element of the FOCUS list, while TOPIC and BG lists are empty:

(10)
$$
\begin{bmatrix} \textit{focus-loc} \\ \text{CTXT} \,|\, \text{INFOSTR} \begin{bmatrix} \text{TOPIC} \langle \rangle \\ \text{FOCUS} \langle \boxed{1} \rangle \\ \text{BG} \langle \rangle \end{bmatrix} \\ \text{CONT} \,|\, \text{HOOK} \,|\, \text{INDEX} \boxed{1} \end{bmatrix}
$$

Topic and background, on the other hand, share the property of an empty FOCUS list, and are thus subtypes of a more general configuration of local features called *non-focus-loc*, from which they inherit the empty FOCUS list as shown below.

(11)

$$
\left[
\begin{array}{l}
\textit{non-focus-loc} \\
\text{CTXT} \,|\, \text{INFOSTR} \,|\, \text{FOCUS} \,\langle\rangle
\end{array}
\right]
$$

$$
\left[
\begin{array}{l}
\textit{topic-loc} \\[4pt]
\text{CTXT} \,|\, \text{INFOSTR} \left[\begin{array}{l}\text{TOPIC } \langle\boxed{1}\rangle \\ \text{FOCUS } \langle\rangle \\ \text{BG } \langle\rangle\end{array}\right] \\[12pt]
\text{CONT} \,|\, \text{HOOK} \,|\, \text{INDEX } \boxed{1}
\end{array}
\right]
\quad
\left[
\begin{array}{l}
\textit{bg-loc} \\[4pt]
\text{CTXT} \,|\, \text{INFOSTR} \left[\begin{array}{l}\text{TOPIC } \langle\rangle \\ \text{FOCUS } \langle\rangle \\ \text{BG } \langle\boxed{1}\rangle\end{array}\right] \\[12pt]
\text{CONT} \,|\, \text{HOOK} \,|\, \text{INDEX } \boxed{1}
\end{array}
\right]
$$

The local types are used in phrasal signs for topical, focused and background constituents. For example, a focused phrase will be defined as a phrase with local features of type *focus-loc*, with the additional constraint that it should bear main stress[3]:

(12)
$$
\left[
\begin{array}{l}
\textit{focus-phrase} \\
\text{STRESS } \textit{main} \\
\text{SYNSEM} \,|\, \text{LOC } \textit{focus-loc}
\end{array}
\right]
$$

Let us now turn to non-focused constituents, starting with the topic. It is generally accepted that definiteness is an important characteristic of the topic, so it should be a constraint on a topic constituent:

(13)
$$
\left[
\begin{array}{l}
\textit{topic-phrase} \\
\text{SYNSEM} \,|\, \text{LOC} \left[\begin{array}{l}\textit{topic-loc} \\ \text{CAT} \,|\, \text{HEAD} \,|\, \text{DEF } \textit{true}\end{array}\right]
\end{array}
\right]
$$

The type *topic-phrase* unifies with definite phrases, in other words definite nominal phrases as well as pronouns. Note that definiteness may be a necessary condition for a linguistic phrase to be considered a topic, but it is obviously not a sufficient condition. In other words, whilst the *topic-phrase* type is used in topic-comment constructions to ensure that the topic is definite, this does not mean that definite phrases are automatically analysed as topics.

Finally, different types of background constituents can be accounted for, for instance:

(14)
$$
\left[
\begin{array}{l}
\textit{bg-adv-phrase} \\
\text{SYNSEM} \,|\, \text{LOC} \left[\begin{array}{l}\textit{bg-loc} \\ \text{CAT} \,|\, \text{HEAD } \textit{adverb-prep}\end{array}\right]
\end{array}
\right]
$$

The type *bg-adv-phrase* applies to adverbials and prepositional phrases[4], and analyses them as background phrases. There is no constraint on the STRESS value of the adverbial since, depending on its position in the sentence, a background adverbial may or may not bear an accent.

(15) a. [F Din ′mor ′ringede] [BG forleden].
       'Your mother CALLED the other day.'
    b. [BG For′leden] ringede telefonen flere gange om natten.
       'The other day, the phone rang several times at night.'

When it is postponed to the focus, as in (15a), a background adverbial is unstressed, whilst it is stressed when placed sentence-initially as in (15b). Therefore, the requirement on accentuation cannot be directly associated with a generic *bg-adv-phrase* type, but must be expressed at a different level in the constituent structure of the sentence. I will return to this point. To complete this short account of the information structure of sentence adverbials, the following sentence is an example of a stressed final adverb which will unify with (12) above to yield focus interpretation.

---

[3] As noted earlier, non-pronominal focused constituents must be stressed in Danish. The opposite does not necessarily hold, i.e. stressed constituents may be part of the background.

[4] The type *adverb-prep* is a supertype of *adverb* and *preposition*.

(16) Hvornår ringede hun? [$_{BG}$ Det var] [$_F$ for'leden].
    'When did she call? It was the other DAY.'

The approach presented here simplifies the issue of how focusing works for adverbials in several respects. Particularly, it does not account for the fact that semantic content also seems to contribute to the information structure function that an adverbial is perceived to have. For example, attitudinal disjuncts such as *måske, desværre* 'perhaps, unfortunately' are not focused in the normal case, while manner adverbials usually are [12]. The issue will be left open here.

5 Information structure principles

It is often assumed that the instantiation of the information structure of the main sentence is achieved by the combination of two general principles called *inheritance* and *projection*. Engdahl and Vallduví express these principles informally in the following way:

> INFO-STRUCT *instantiation principles for English*:
> *Either* (i) if a DAUGHTER's INFO-STRUCT is instantiated, then the mother inherits this instantiation (*for narrow foci, links and tails*),
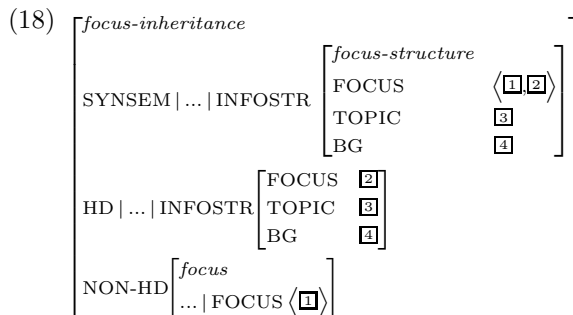> *or* (ii) if the most oblique DAUGHTER's FOCUS is instantiated, then the FOCUS of the mother is the sign itself (*wide focus*)
> (Engdahl and Vallduví 1996:12)

The former principle corresponds to inheritance, the latter to focus projection. The focus projection principle, in particular, predicts a wide focus structure in cases where sentence accent falls on the most oblique verbal complement. I have already noted, however, that there is no sentence accent in Danish[5], and that each verbal complement or adjunct in the focus domain is accented. Therefore, it is not clear that the focus projection principle should play a role in Danish, and here I will only give a formal account of information structure inheritance. What we need to express is the fact that for each information structure feature, the value of the mother node is a list consisting of the corresponding values of the two daughters. Engdahl and Vallduví argue convincingly that such an approach, in which information structure values are collected from the daughters of a node without reference to syntactic constituency, is an advantage in that it can deal with cases in which there is no direct correspondance between these different aspects of a linguistic sign. Consider the following Danish example:

(17) Hvor blev mine lygter af? Dem [$_F$ TOG THOMAS].
    Where disappeared my lights off? Them took Thomas.
    'Where have my bicycle lights disappeared? Thomas took them.'

The focused part of the sentence does not form a syntactic constituent, however it is a unit from an information structure point of view. This is not a problem for an approach where focus values are simply collected from daughters to mother node. Information structure inheritance is expressed here by means of a number of specific inheritance principles. The principle of *focus-inheritance* is shown below:

(18)
$$
\begin{bmatrix}
\textit{focus-inheritance} \\
\text{SYNSEM}\,|\dots|\,\text{INFOSTR}
\begin{bmatrix}
\textit{focus-structure} \\
\text{FOCUS} & \left\langle \boxed{1},\boxed{2} \right\rangle \\
\text{TOPIC} & \boxed{3} \\
\text{BG} & \boxed{4}
\end{bmatrix} \\
\text{HD}\,|\dots|\,\text{INFOSTR}
\begin{bmatrix}
\text{FOCUS} & \boxed{2} \\
\text{TOPIC} & \boxed{3} \\
\text{BG} & \boxed{4}
\end{bmatrix} \\
\text{NON-HD}
\begin{bmatrix}
\textit{focus} \\
\dots|\,\text{FOCUS} \left\langle \boxed{1} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

---

[5] Unless the sentence displays particular emphasis or a contrastive focus.

The principle states that in a binary projection in which the non-head daughter is a focused constituent, the focus value of the non-head is appended to the focus value of the head daughter (which may or may not contain additional focused elements), and the resulting list is the FOCUS value of the mother. TOPIC and BG values are copied from head to mother node. This ensures the accumulation of focus values on the top node of the syntactic tree as well as inheritance of the other information status features. The application of focus, topic and background inheritance is shown in Fig. (4)[6].

S

$$\begin{bmatrix} \textit{topic-inheritance} \\ \text{TOPIC} \; \langle \boxed{1} \rangle \\ \text{FOCUS} \; \langle \boxed{3}, \boxed{2} \rangle \\ \text{BG} \; \langle \boxed{4} \rangle \end{bmatrix}$$

NP

$$\begin{bmatrix} \text{TOPIC} \; \langle \boxed{1} \rangle \\ \text{DEF true} \end{bmatrix}$$

VP

$$\begin{bmatrix} \textit{bg-inheritance} \\ \text{TOPIC} \; \langle \rangle \\ \text{FOCUS} \; \langle \boxed{3}, \boxed{2} \rangle \\ \text{BG} \; \langle \boxed{4} \rangle \end{bmatrix}$$

VP

$$\begin{bmatrix} \textit{focus-inheritance} \\ \text{FOCUS} \; \langle \boxed{3}, \boxed{2} \rangle \\ \text{BG} \; \langle \rangle \end{bmatrix}$$

ADVP

$$\begin{bmatrix} \text{BG} \; \langle \boxed{4} \rangle \\ \text{STRESS none} \end{bmatrix}$$

V

$$\begin{bmatrix} \text{FOCUS} \; \langle \boxed{2} \rangle \\ \text{STRESS main} \end{bmatrix}$$

ADVP

$$\begin{bmatrix} \text{FOCUS} \; \langle \boxed{3} \rangle \\ \text{STRESS main} \end{bmatrix}$$

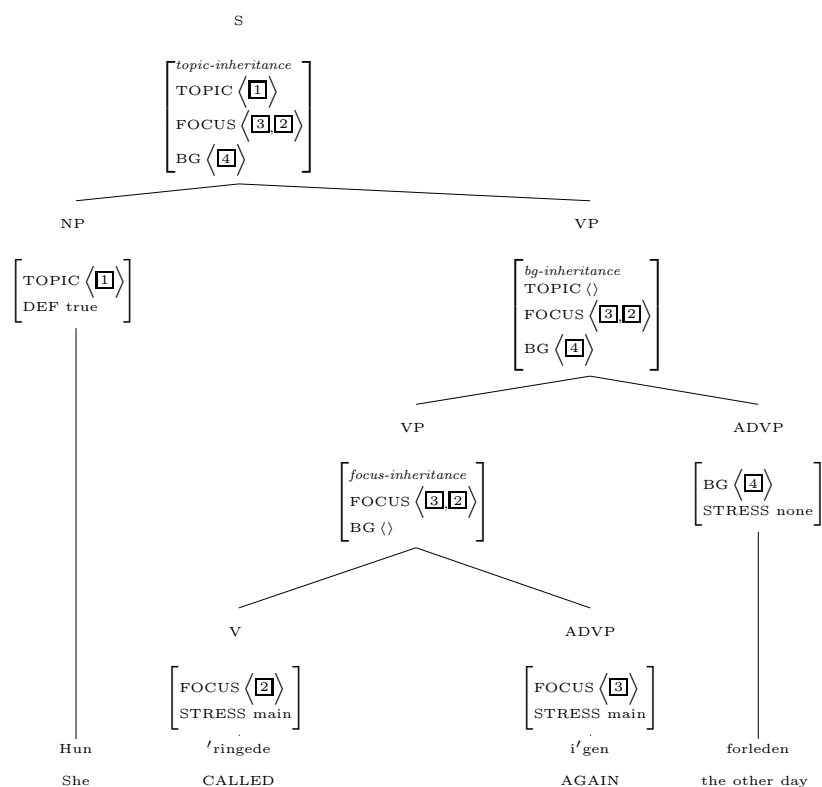| Hun	'ringede	i'gen	forleden
She	CALLED	AGAIN	the other day

Fig. 4. Information structure principles at work

The background and topic inheritance principles are very similar to focus inheritance: they ensure that either the background or the topic value of a non-head daughter is appended to the background value of the head, while the remaining information structure features are simply passed on from head daughter to mother.

The figure shows the analysis of the following sentence:

(19) Hun 'ringede 'igen forleden.
'She called again the other day.'

In the most obvious reading, this is a topic-comment construction. The initial pronoun is the topic, and unifies with the *topic* type in that it is a personal pronoun and therefore definite. Both the verb and the adverb adjacent to the verb are stressed, and therefore compatible with a focus analysis, whilst the unstressed final adverbial must be outside of the focus domain. For each binary projection, it is indicated in the feature structure type which principle has applied.

---

[6] The feature paths are simplified for the sake of space. For instance, TOPIC should be read SYNSEM|LOC|CTXT|INFOSTR|TOPIC.

Note that the non-head daughter in a phrase to which focus inheritance applies is constrained to being a *focus* phrase, defined in (12). Focus inheritance will therefore not apply to background constituents, and wrong analyses like the following one will not be produced:

(20) *[$_F$ Hun ′ringede ′igen forleden].
    'Your mother CALLED the other day.'

6 Construction types

The information structure types and principles discussed so far come into play in different phrasal and clausal types to give rise to what I call information structure constructions.

6.1 Phrasal constructions and information structure

Within HPSG, constructions are modelled in terms of phrasal types that are cross-classified along several dimensions, the two most important ones being *clausality* and *headedness* [13] [14]. Multiple inheritance ensures that the features of the more general types are also shared by the more specific ones. Such an approach makes it possible to define clausal as well as phrasal constructions, both of which I believe interact with the information structure principles and types presented earlier. This interaction is shown in Fig. (5) for what concerns head-complement (*hd-comp*) and head-adjunct (*hd-adj*) projections. The resulting types take care of information structure inheritance – whether focus or background inheritance – within the VP.
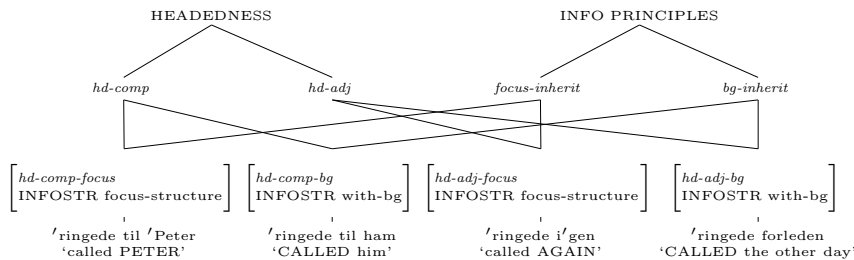

Fig. 5. Information structure of complements and adjuncts

A similar cross-classification is posed for head-subject (*hd-subj*) and filler-head (*filler-hd*) types to ensure that information structure values are assigned at the level of the top sentence node, whether a declarative or a topicalised sentence.

I argued earlier in the discussion of the examples in (15) that background adverbials in Danish are unstressed when placed at the end of the sentence, and stressed when they occur at the beginning. The complex phrasal constructions that result from the unification of phrasal types and information structure principles are the place where such a constraint can be formalised. The fact that a postverbal background constituent should be unstressed can be expressed in the type that combines the *hd-adj* phrase with the *bg-inherit* principle. The non-head daughter of this type unifies, in fact, with a postverbal adverbial or propositional phrase.

(21) $\begin{bmatrix} \textit{hd-adj-bg} \\ \text{NON-HD}\begin{bmatrix} \text{STRESS false,} \\ \text{SYNSEM} \mid \text{CAT} \mid \text{HEAD } \textit{adverb-prep} \end{bmatrix} \end{bmatrix}$

On the other hand, the constraint that background adverbials in initial sentence position must bear an accent is expressed in a type combining *filler-hd* and *bg-inherit*. In our grammar, in fact, an initial sentence adverbial unifies with the non-head daughter of a *filler-hd* phrase.

(22) $\begin{bmatrix} \textit{filler-hd-bg} \\ \text{NON-HD} \begin{bmatrix} \text{STRESS true,} \\ \text{SYNSEM | CAT | HEAD } \textit{adverb-prep} \end{bmatrix} \end{bmatrix}$

6.2   Clausal constructions and information structure

Let us now turn to clausal constructions and see how information structure properties are realised in them. Following the HPSG view of grammatical constructions, in [15] Anne Neville and I have defined a typology of Danish clause types which accounts for the different word orders and extraction patterns allowed in the various types. The most general (leaving out in particular relative clauses) are shown in Fig. (6).
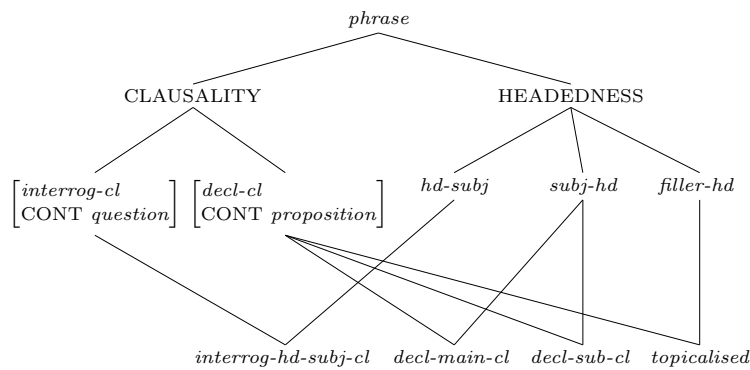
Fig. 6. Clausal constructions

The type *interrog-hd-subj-cl* accounts for questions, *decl-main-cl* for declarative main clauses, *decl-sub-cl* for declarative subordinate clauses, and *topicalised* for topicalised constructions. In order for the various types of information structure to become part of the definition of clausal construction types, the type hierarchy in Fig. (6) must be enriched so that the most specific types are cross-classified with respect to an information structure dimension. In Fig. (7) we can see how this approach can account for the interaction between the topic-comment information structure and two types of clausal constructions, i.e. main declarative and topicalised sentences.
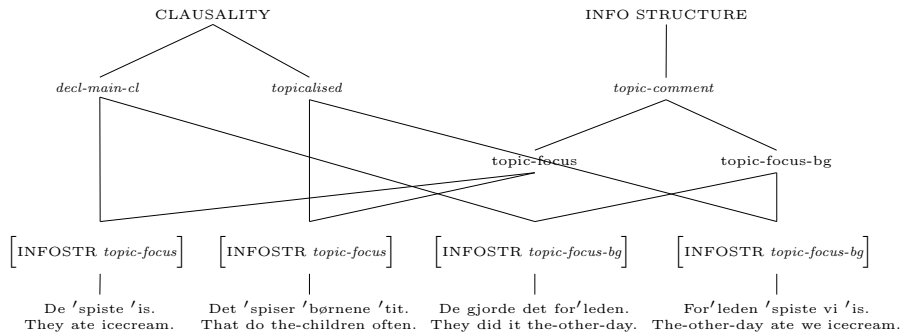

Fig. 7. Information structure in declarative main clauses

Four different specific constructions are modelled:

(23) a. $[_T$ De] $[_F$ ′spiste ′is].
         'They ate icecream.'
     b. $[_T$ Det] $[_F$ ′spiser ′børnene ′tit].
         That do the-children often.
         'The children often do that.'
     c. $[_T$ De] $[_{BG}$ gjorde det] $[_F$ for′leden].
         They did it the-other-day.
         'They did it the other day.'
     d. $[_{BG}$ For′leden] $[_F$ ′spiste] $[_T$ vi] $[_F$ ′is].
         The-other-day ate we icecream.
         'The other day they ate icecream.'

All of them are topic-comment constructions as expected. However, the topic has different grammatical functions and is placed in different positions. In the last example it is actually interposed between two focused constituents. As already mentioned, the fact that information structure values are not homomorphic with syntactic constituents is not a problem in an approach where these values are collected in lists for each of the binary projections in the syntactic tree.

Topicless constructions are modelled along the same lines by letting clausal types be cross-classified with the two topicless types *all-focus* and *focus-bg*, and specific constructions such as clefts or existential sentences can be formalised in that part of the construction hierarchy.

7   Conclusion

To sum up, in this paper I have presented a proposal for the representation of the information structure of a sentence by way of features in a typed feature structure formalism.

The interaction between information structure and grammar constructions has been formalised in a hierarchy where different phrasal and clausal types of Danish grammar are cross-classified with respect to an information structure dimension. Examples of the complex types that derive from this cross-classification are topic-comment constructions expressed as either main declarative clauses or topicalised sentences. I have also shown that the feature types defined in the hierarchy can account for prosodic and syntactic constraints, for instance the different stress patterns displayed by background adverbials in initial or final position, or the definiteness of sentence topic.

I plan to extend the model to include an account of syntactically marked topicless sentences such as clefts and existential sentences, and to investigate the way in which the information structure principles and types work in unbounded dependencies, of which topicalised constructions and clefts are examples.

References

1. Paggio, P.: The treatment of information structure in machine translation. PhD thesis, University of Copenhagen and Centre for Language Technology (1997).
2. Paggio, P.: The information structure of Danish grammar constructions. Constructions (Under publication).
3. Pollard, C., Sag, I.A.: Head-Driven Phrase Structure Grammar. Chicago: The University of Chicago Press (1994).
4. Copestake, A., Flickinger, D., Pollard, C., Sag, I.A.: Minimal recursion semantics: An introduction. Manuscript: available at www-csli.stanford.edu/ aac/papers/newmrs.pdf (1999).
5. Flickinger, D., Bender, E.M.: Compositional semantics in a multilingual grammar resource. In: Proceedings of the Workshop on Ideas and Strategies for Multimodal Grammar Development, ESSLLI (2003) 33–42.
6. Engdahl, E., Vallduví, E.: Information packaging in HPSG. In Grover, C., Vallduví, E., eds.: Studies in HPSG. Volume 123 of Edinburgh Working Papers in Cognitive Science. Centre for Cognitive Science, University of Edinburgh (1996) 1–32.
7. Alexopoulu, T., Kolliakou, D.: On linkhood, topicalization and clitic left dislocation. Linguistics **38** (2002) 193–245.

8. Engdahl, E.: Integrating pragmatics into the grammar. In Mereu, L., ed.: Boundaries of Morphology and Syntax. John Benjamins: Amsterdam (1999) 175–194.
9. Flickinger, D., Bender, E.M., Oepen, S.: MRS in the LinGO Grammar Matrix: A Practical User's Guide (2003).
10. Cooper, R.: Using type theory with records for hpsg. Invited talk, Fest Colloquium for Uwe Mnnich, Freudenstadt (2004).
11. Cann, R.: Formal Semantics. Cambridge Textbooks in Linguistics. Cambridge University Press (1993).
12. Togeby, O.: PRAXT. Pragmatisk tekstteori. Århus: Aarhus Universitetsforlag (1992).
13. Sag, I.: Aspects of a theory of grammatical constructions. Lecture given at the First International Construction Grammar Conference, Berkeley (2001).
14. Ginzburg, J., Sag, I.A.: Interrogative Investigations: Form, Meaning and Use of English Interrogatives. CSLI Publications, University of Chicago Press (2000).
15. Neville, A., Paggio, P.: Developing a Danish grammar in the GRASP project: a construction-based approach to topology and extraction in Danish. Forthcoming in Electronic Notes in Theoretical Computer Science 47 (2001).