Extraction from Relative and Embedded Interrogative Clauses in Danish

Anne Bjerre
University of Southern Denmark
Engstien 1
DK-6000 Kolding
bjerre@sitkom.sdu.dk

Abstract
In Danish relative clauses and embedded interrogative clauses are not extraction islands. However, there is an asymmetry between the two clauses. In Danish it is possible to extract the subject out of an embedded interrogative clause. Extraction of the subject out of a relative clause, on the other hand, is not allowed. In this paper we present a formal HPSG analysis of extraction in Danish which treats the extraction out of relative and embedded interrogative clauses in a uniform manner, and the asymmetry between the clauses will be shown to follow from a more general constraint on adjuncts.

1 Introduction
Relative clauses and embedded interrogative clauses are commonly taken not to allow extraction. This has been extensively discussed under the headings of extraction islands, (Ross, 1967), subjacency and the Empty Category Principle, (Chomsky, 1973). In Danish, relative clauses and embedded interrogative clauses are not extraction islands. However, there is an asymmetry between the two clauses, as also noted by e.g. Engdahl (1984), in that it is not possible to extract the highest subject out of a relative clause, a restriction that does not apply to embedded interrogatives.

It has been suggested that there is a structural difference between relative clauses and embedded interrogative clauses that might explain the different behaviour when it comes to subject extraction, e.g. Engdahl (1984). She suggests that relative clauses are of category $\mathcal{S}$, whereas interrogative clauses are of type $\mathcal{\exists}$ which has an extra XP position. This difference has the consequence that the empty category in subject position in relative clauses is not properly governed. Consequently the Empty Category Principle rules out extractions of subjects from relative clauses.

In feature-based analyses it has been suggested that subject extraction does not involve extraction at all, e.g. Gazdar (1981) and Pollard and Sag (1994). However, in more recent feature-based analyses, e.g. Sag (1997), Bouma et al. (2000), Ginzburg and Sag (2000) and Sag (to appear), subjects are also treated as being extracted. The analyses have been influenced by discussions by e.g. Hukari and Levine (1996) concerning subject extraction, arguing that there is cross-language evidence that subjects are also extracted in the same way as complements.

In this paper we provide a feature-based analysis of extraction that treats the extraction of subjects out of embedded interrogatives and relative clauses in a uniform manner, and show that the asymmetry between the clauses follow from a more general constraint on adjuncts clauses.

In section 2 we show the relevant Danish extraction data. Then in section 3 we go on to present the analytical background that the formalization of our analysis is based on. Then our analysis is formalized in section 4. In section 5 we discuss the Complex NP Constraint, and finally we conclude in section 6.

2 The Danish data
As stated above, relative and embedded interrogative clauses are not extraction islands in Danish. In (1) we find examples of object extractions from these clauses.¹

(1) a. Bøden fatter jeg ikke hvem
The fine understand I not whom
there has received

¹The examples in the present paper are from Hansen (1974), KORPUSDK and the Internet.
Extraction from relative and embedded interrogative clauses in Danish

b. Betændelse kender jeg mange der
   Infection know I many there
   have had, in a mild form

In (2) we find examples of subject extractions from embedded interrogatives.

(2) a. Nu har jeg fundet det dokument,
   Now have I found the document,
   som jeg i går ikke vidste hvor
   which I yesterday not knew where
   var.
   was

b. *Jeg traf en fyr som jeg bare ikke
   I met a guy I just not
   kan huske hvor der boede.
   can remember where there lived

In contrast, (3) shows that subject extraction out of relative clauses is not possible.

(3) a. Tv-nævnet bad TV2 om en
   The Tv committee asked TV2 for a
   redegørelse i sagen, hvilken
   report in the case, which
   Pernille fik til opgave at skrive.
   Pernille got to task to write
   ’The Tv committee asked TV2 for a
   report in the case, which Pernille got
   the task of writing’

b. *Pernille bad tv-nævnet
   Pernille asked The Tv committee
   TV2 om en redegørelse i sagen,
   TV2 for a report in the case, which
   got to task to write
   hvilken fik til opgave at skrive.
   got to task to write

c. *Pernille bad tv-nævnet
   Pernille asked The Tv committee
   TV2 om en redegørelse i sagen,
   TV2 for a report in the case, which
   got to task to write
   hvilken der fik til opgave at skrive.
   which got to task to write

d. Han stillede et spørgsmål, som
   He asked a question, C
   ministeren ikke var forberedt på.
   the minister not was prepared for

e. *Ministeren stillede han et
   The minister asked he a
   spørgsmål, som ikke var forberedt
   question, C not was prepared

f. *Ministeren stillede han et
   The minister asked he a
   spørgsmål, som der ikke var
   question, C there not was
   forberedt på.
   prepared for

Apart from the difference in extraction potential between subjects and objects, the insertion of expletive der (‘there’) instead of a gap in some of the clauses should be noted. As observed in Hansen (1974), der is inserted in local subject extractions in standard Danish, cf. (1a) and (1b). In non-local subject extractions der is not inserted in standard Danish, cf. (2a). In non-standard Danish der insertion is also found in non-local subject extractions, cf. (2b).

3 Analytical background

Before our analysis is presented we will go through the analytical background that we base our analysis on. The analysis is largely based on the feature-based account of extraction in Ginzburg and Sag (2000), cf. also Neville and Paggio (2004) for an analysis of Danish relative clause constructions. The presentation of the formal background here is by no means exhaustive, and the semantics of the analysis is left out. However, a number of important assumptions for the purpose of understanding the formalization in section 4 are explained. The grammar used is a hierarchy of typed feature structures with associated implicational constraints constraining what constitutes well-formed linguistic entities.

Within the HPSG framework, constructions involving extractions are called filler-gap constructions. Filler-gap constructions are specified for the feature SLASH, and filler-gap dependencies are established through the inheritance of SLASH specifications. A non-empty SLASH specification is introduced at the lowest level where the “gap” is introduced, then passed up through the structure, to be bound off by the filler. In (5) the specification for the SLASH feature in a filler-gap construction, like the highlighted part of (4), is illustrated.

(4) Han var klippen på hvilken, man stod.
   He was the rock on which one stood
A “gap” in the structure is a feature structure of type \( \text{gap-ss}(yn)stem \). The \( \text{gap-ss} \) type does not have any phonological content, and it shares its syntactic and semantic content with its corresponding overt element. It is constrained by the constraint in (6).

\[
\text{(6) } \text{gap-ss} \implies \begin{array}{c}
\text{LOC } \Box \\
\text{SLASH } \{\Box\}
\end{array}
\]

The structure in (5) is constrained by a set of constraints. The non-empty SLASH feature is introduced by the Argument Realization Principle, (Ginzburg and Sag, 2000, p. 171), and the SLASH-Amalgamation Constraint, (Ginzburg and Sag, 2000, p. 169). These are shown in (7) and (8), respectively.

\[
\text{(7) Argument Realization Principle: }
\begin{array}{c}
\text{word} \implies \\
\text{SS } \mid \text{LOC } \mid \text{CAT} \\
\text{\text{SUBJ } } \exists \\
\text{\text{COMPS } } \Box \ominus \text{list(gap-ss)} \\
\text{ARG-ST } \Box \ominus \Box \ominus \Box
\end{array}
\]

\[
\text{(8) SLASH-Amalgamation Constraint: }
\begin{array}{c}
\text{word} \implies \\
\text{SS } \mid \text{SLASH } \Box \ominus \text{... } \ominus \text{.. } \\
\text{\text{ARG-ST } } \Box \ominus \text{... } \ominus \text{.. } [\text{SLASH } \Box \ominus]
\end{array}
\]

The Argument Realization Principle may introduce a “gap” on the ARG-ST list of a word, at the same time, removing \textit{synsems} that have been resolved to \textit{gap-ss} from the COMPS list of a word. The SLASH-Amalgamation Constraint ensures that the SLASH values of the arguments of a word are inherited by the word itself. This is also known as lexicalized SLASH-amalgamation. The SLASH-Amalgamation Constraint is a default constraint. The inheritance of the SLASH value in constructions is taken care of by the Generalized Head Feature Principle, (Ginzburg and Sag, 2000, p. 33), which specifies inter alia the inheritance of the SLASH feature from the head-daughter to the mother in a construction.

Finally, various constraints are responsible for binding off the SLASH value, either constraints involving a filler daughter or constraints involving constructional gap-binding. Constraints involving a filler daughter are constraints that are subtypes of the more general constraint on \textit{hd-fill-ph}, e.g. \textit{wh}\textsuperscript{2}-interrogative clauses. This constraint is shown in (9), cf. Ginzburg and Sag (2000, p. 174).

\[
\text{(9) } \text{hd-fill-ph: }
\begin{array}{c}
\text{SLASH } \Box \ominus \\
\text{phrase} \\
\text{LOC } \Box, \text{H} \\
\text{HEAD } v \\
\text{SLASH } \{\Box\} \cup \{\Box\}
\end{array}
\]

This constraint removes the \textit{gap-ss} from the SLASH set which corresponds to the \textit{LOCAL} value of the filler daughter.

Constructional gap-binding occurs in e.g. relative clauses which do not contain relative pronouns. In such cases a certain construction binds off the “gap” instead of a \textit{wh-word}. Sag (1997) introduces the \textit{non-wh-rel-cl} with the constraint shown in (10).

\[
\text{(10) } \text{non-wh-rel-cl: }
\begin{array}{c}
\text{HEAD } \mid \text{MOD } \tilde{N}
\end{array} \rightarrow \text{H} \left[\text{SLASH } \{\text{NP}\} \right] \\
\text{SLASH } \{\}
\]

The effect is to build a unary branching structure which turns a clause into a relative clause by binding off the “gap”, while at the same time introducing a \textit{MOD} feature and co-indexing the index on the element in the SLASH set with the index of the \textit{MOD} value. In this way co-indexing between the modified noun and the “gap” is ensured. The

\footnote{We use “\textit{wh}” for the Danish “hv” words.}
highlighted part of (11) is an example of constructional gap-binding which is shown in (12).³,⁴

(11) Det er fordi de i naturen skal
   This is because they in the nature must
   kunne følge de fisk, som de spiser.
   could follow those fish, they eat

(12) FORM(\{som, de spiser\})
    \[\text{LOC CP,\{SUBJ\}} \mid \text{HEAD MOD \{\}} \]
    SOLASH(\{\})

FORM(\{som, de spiser\})
    \[\text{LOC CP,\{SUBJ\}} \mid \text{HEAD MOD \{\}} \]
    SOLASH(\{\})

FORM(\{som\})
    \[\text{LOC C,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

FORM(\{de spiser\})
    \[\text{LOC V,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

FORM(\{de\})
    \[\text{LOC C,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

FORM(\{spiser\})
    \[\text{LOC V,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

FORM(\{\})
    \[\text{LOC C,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

FORM(\{\})
    \[\text{LOC C,\{} \mid \text{COMP(S)} \}
    SOLASH(\{\})

Note that because of the Generalized Head Feature Principle, the SOLASH value is passed up to the clause via the complementizer which is analyzed as the head of the clause.

4 Formal analysis

In this section the proposed formal analysis of Danish extraction is presented. The analysis is a further development of the analysis presented in Bjerre (to appear).

As explained in section 3, a “gap” in the structure is a feature structure of type gap-s(yn)s(em) which does not have any phonological content, cf. the constraint in (6). To account for the Danish der insertion phenomenon in certain subject extraction contexts, we introduce another type of non-canonical synsem type. The extended hierarchy is shown in (13).

(13) synsem
canón-ss noncan-ss
expl-ss gap-ss pró-ss

The type expl-ss is a synsem which has phonological content, and it is consequently a subtype of canon-ss. But, as the gap-ss, the expl-ss does not have any semantic content of its own. The two latter types differ in that the gap-ss also has no syntactic content, unlike the expletive. (14) shows the constraint for expl-ss.

(14) expl-ss =⇒ [LOC [CAT | HEAD expl]
    [SOLASH \{\{\}] \]}]

In Danish, the Argument Realization Principle additionally removes synsems that have been resolved to gap-ss from the SUBJ list of the word. However, the subject is visible as the value of the SUBJECT feature.⁵ The Danish Argument Realization Principle is shown in (15).

(15) Argument Realization Principle (Danish):

word =⇒

³We follow Erteschik-Shir (1984) and Vikner (1991) in treating “som” as a complementizer. This is motivated e.g. by (non-standard) examples like (1) where we have combinations of wh-pronouns and “som”, and another complementizer “at” (‘that’). In such cases, the wh-pronoun binds off the slash value.

³Another example is relative clauses without both relative pronoun and “som”, as the example in (1).

⁴The nature Principle, the synsem is a feature structure of type gap-ss.

⁵Cf. Meurers (1999) for further arguments that we need a SUBJECT feature as part of the HEAD feature.
The element in the head daughter’s SLASH set which corresponds to the filler daughter is bound in Danish. If the subject is resolved to an existential realization of an expletive in subject position on the SUBJ it remains on the hd-subj-ph the expletive subject corresponding to an element in the SLASH set. This means that a SLASH value originating from an expletive can only be bound off locally. The Expletive SLASH Constraint applies in standard Danish, but as we saw in section 2, not in non-standard Danish.

Now we come to the binding off of the SLASH value. For the present purpose, extraction out of embedded interrogatives and relative, we need the types listed in (19), cf. the clause hierarchy set up in Ginzburg and Sag (2000).

(19)  

a. fin-wh-ns-int-cl  

b. fin-wh-ns-rel-cl  

c. fin-non-wh-rel-cl

In the following we will concentrate on what constraints are involved in structures of these types, but we will not be specific about where the constraints originate from. Some of the constraints may be specific to these types, others may be inherited from more general types.

The data we need to account for with the fin-wh-ns-int-cl construction are (2) repeated in (20). The construction covers the highlighted parts of the examples.

(20)  

a. Nu har jeg fundet det dokument,  
Now have I found the document,  
som jeg i går ikke vidste hvor  
I didn’t know where  
C was  
vhar  

b. ?Jeg traf en fyr som jeg bare ikke  
I met a guy C I just not  
kun huske hvor der boede.  
can remember where there lived

The constraints for the highlighted embedded wh-interrogative clauses in (20) are given in (21).

(21)  

fin-wh-ns-int-cl

The element in the head daughter’s SLASH set which corresponds to the filler daughter is bound
off. The SLASH set may contain a second element, corresponding to an extracted subject. This element is not bound off, but passed on to the mother. The value of the head daughter’s scshape subject list may in that case be either a gap-ss, as in (20a) or a expl-ss as in (20b). If both the SLASH-Amalgamation Constraint and the Expletive SLASH Constraint are applied, only the example without an expletive subject is well-formed, as an element in the SLASH set corresponding to the expletive subject will not be amalgamated by a selecting head. If only the SLASH-Amalgamation Constraint is applied both examples are well-formed, as the SLASH set corresponding to an expletive subject is not excluded by the Expletive SLASH Constraint. Importantly, nothing prevents a subject from being extracted.

We turn now to the relative clause examples. The data we need to account for with the fin-wh-ns-rel-cl construction are the examples in (3b) and (3c) repeated in (22). The construction covers the highlighted parts.

(22) a. *Pernille bad tv-nævnet
Pernille asked The Tv committee
TV2 om en redegørelse i sagen,
TV2 for a report in the case,
hvilken fik til opgave at skrive.
which got to task to write

b. *Pernille bad tv-nævnet
Pernille asked The Tv committee
TV2 om en redegørelse i sagen,
TV2 for a report in the case,
hvilken der fik til opgave at skrive.
which there got to task to write

We use an extended ARG-ST list for words as in Ginzburg and Sag (2000). There it is used for certain optionally selected adjuncts. We assume that noun words may be derived from noun lexemes with an additional element on the ARG-ST list, i.e., a restrictive relative clause. This means that the Argument Realization Principle will give rise to a representation as shown in (23) with a restrictive relative clause on the COMPS list.

(23) 

The consequence of this analysis is that the SLASH value of a restrictive relative clause will be passed on to the head noun by the SLASH-Amalgamation Constraint.

The important constraint on finite adjunct clauses that gives rise to the asymmetry between relative clauses and embedded interrogative clauses is introduced in (24).

(24) 

The constraint expresses the generalization that adjunct clauses require a subject. The constraint excludes adjunct clauses which contain a non-canonical subject which has not already been bound off.

The fin-wh-ns-rel-cl is a subtype of fin-adjunct-cl and is subject to the constraint in (24). The constraints for non-subject wh-relative clauses is given in (25).

(25) fin-wh-ns-rel-cl

As in (21) the element in the head daughter’s SLASH set which corresponds to the filler daughter is bound off and the SLASH set may contain a second element, corresponding to an extracted subject. The value of the head daughter’s scshape subject list may again be either a gap-ss, as in (22a) or a expl-ss as in (22b). This element is not bound off, but passed on to the mother. If the SLASH-Amalgamation Constraint and the Expletive SLASH Constraint are both applied to the examples in (22), only the example with a gap-ss subject will pass on the corresponding SLASH element. However, because of the constraint in (24) on adjunct clauses, the example in (22a) is
ill-formed, as it contains a non-canonical subject not bound off. If only the SLASH-Amalagamation Constraint is applied, the example with a expl-ss subject will also pass on the corresponding SLASH element. However, again because of the additional contraint in (24) on adjunct clauses, the example in (22b) is also ill-formed. Importantly, in neither case is it allowed to extract the subject.

With respect to the fin-non-wh-rel-cl, it is the examples in (3e) and (3f) repeated in (26) we need to exclude. The construction covers the highlighted part of the examples.

(26)  a. *Ministeren stillede han et
    The minister asked he a
    spørgsmål, som ikke var forberedt
    question, C not was prepared
    på.
    for
    b. *Ministeren stillede han et
    The minister asked he a
    spørgsmål, som der ikke var
    question, C there not was
    forberedt på.
    prepared for

The constraints on the fin-non-wh-rel-cl are shown in (27).

(27)  non-wh-rel-cl:

\[
\begin{array}{c}
\text{SS} \\
\text{LOC} \mid \text{CAT} \\
\text{SLASH} \\
\text{HEAD} \{ \text{MOD} \} \\
\text{SUBJ} \{ \} \\
\end{array}
\]

In (27) the “object” element in the head daughter’s SLASH set is bound off constructionally. Again the SLASH set may contain a second element, corresponding to an extracted subject which is passed on to the mother. Again the value of the head daughter’s subject list may be either a gap-ss, as in (26a) or a expl-ss as in (26b). In any case, because of the constraint on finite adjunct clauses, a SLASH element corresponding to the highest subject in the clause will not be passed on for the noun to amalgamate so that it can escape the noun phrase. Again, we see it is not allowed to extract a subject in (26a) and (26b) as predicted.

5 The Complex NP Constraint

We have already seen that Ross’ Complex NP Constraint does not apply for Danish NPs with relative clauses, when we looked at object extraction from relative clauses. The constraint that finite adjunct clauses cannot contain a non-canonical subject which is not already bound off rules out subject extraction from relative clauses in an NP. Now, our analysis predicts that subject extraction out of complement clauses in NPs is allowed, and hence that the Complex NP Constraint does not apply in Danish. The examples in (28) contain subject extractions out of complement clauses in NPs.

(28)  a. Teltet er jeg ikke i tvivl
    The tent am I not in doubt
    om ville være et hit.
    whether would be a hit
    b. Markedet er afventende forud for
    The market is hesitant before
    eftermiddagens længere ventede
    the afternoon’s long awaited
    amerikanske arbejdsmarkedssrapport
    American labour market report
    for februar, som der har været
    for February, C there have been
    spekulationer om vil overraske
    speculations whether will surprise
    positivt.
    positively
    c. For så bliver der læst
    Because then is there put
    noget ind i mine ord, jeg
    something in to my words, I
    ikke selv har haft intentioner om
    not self have had intentions about
    skulle være der.
    should be there
    d. Et samarbejde hun udtrykte et
    A cooperation she expressed a
    dybtfølt håb om vil fortsætte.
    deepfelt hope about will continue

Our analysis thus correctly predicts that Danish does not adhere to the Complex NP constraint, cf. also Allwood (1976) and Andersson (1982) for a discussion of the Complex NP Contstraint in Swedish.
6 Conclusion and further research

The paper has shown that Danish allows extraction extensively. We have presented a formal HPSG analysis of extraction which treats the extraction out of relative and embedded interrogative clauses in a uniform manner. The apparent asymmetry wrt. extraction between the two clause types have been shown to follow from a more general constraint on adjunct clauses, i.e. Danish finite adjunct clauses require a subject. If the subject is non-canonical, it must have be bound off internally. We have presented data which show that subject extraction out of complement clauses in NPs is possible. This was shown not to contradict our analysis, rather it follows naturally as our constraint on adjunct clauses does not apply to NP complements. Finally it should be noted that the apparent lack of syntactic constraints on extraction, with varying degrees of acceptability, admittedly, suggests that further research into the pragmatics of extraction is called for in Danish, cf. Erteschik-Shir (1973) and Erteschik-Shir (1982), and for Swedish Allwood (1976) and Engdahl (1997).

References

Jens Allwood. 1976. The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. University of Massachusetts Occasionally Papers in Linguistics II.


