



GÖTEBORGS UNIVERSITET  
INST FÖR SVENSKA SPRÅKET

GU-ISS-2013-02

# Strange things happen on extraction paths

Elisabet Engdahl



**Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet**  
**Research Reports from the Department of Swedish**

**ISSN 1401-5919**

[www.svenska.gu.se/publikationer/GU-ISS](http://www.svenska.gu.se/publikationer/GU-ISS)



## Strange things happen on extraction paths

Elisabet Engdahl

[elisabet.engdahl@svenska.gu.se](mailto:elisabet.engdahl@svenska.gu.se)

My point of departure is a lecture given in the spring of 1981 at Stanford by Ivan Sag and Gerald Gazdar, using a major technical innovation: They had one overhead projector each, and projected in tandem. The lecture was on GPSG and the major theoretical innovation using SLASH categories. This meant a shift of focus from movement transformations to encoding the information as part of a complex syntactic category. Gazdar's original notation is shown in (1a). In HPSG, the SLASH feature is included in the feature structure, as outlined in (1b), involving unification between the initial constituent, the filler, and the bottom of the dependency.

(1) a.  $S \rightarrow NP S/NP$  (Gazdar 1981)

b.  $S \rightarrow_{[1]} NP S_{SLASH [1]}$

The SLASH feature is instantiated on the path which connects the filler and the gap, the so called extraction path,

The notion of extraction path has proved to be a useful analytic tool for a number of phenomena such as expletive suppression in Icelandic, complementizer selection in Irish, tonal downstep in Kikuyu (see Zaenen 1983 for an overview) and of course parasitic gaps (Engdahl 1985, Kayne 1984). The topic of this paper is an on-going change in Swedish which looks rather puzzling at first, but which becomes understandable in terms of extraction paths.

Swedish is a V2-language; the finite verb appears as the second constituent in declarative clauses. Over the years Swedish has developed a strong subject requirement; all tensed clauses require an overt subject. If there is no referential subject, an expletive is used, normally the third person singular personal pronoun *det* ('it'). The expletive subject is used with weather predicates, in various impersonal constructions and in sentences involving extraposed clauses, both finite and infinitival. An example with an extraposed *att*-clause is shown in (2). For perspicuity I mark expletive uses as *det<sub>x</sub>*.

(2) a. (Det<sub>x</sub>) var bra att du kunde komma.  
*it was nice that you could come*

b. Då var \*(det<sub>x</sub>) bra att du kunde komma.  
*so was it nice that you could come*  
'So it was nice that you could come.'

As shown in (2), the expletive can be omitted in the initial, preverbal, position (often referred to as Spec CP), but not in the postverbal position (Spec IP). Note that this pattern is opposite to the one found in Icelandic where expletive *það* only occurs clause initially.<sup>1</sup>

Against this background, it is rather surprising that people say things like in (3a).<sup>2</sup>

- (3) a. Det<sub>i</sub> var dumt att du sa <sub>-i</sub>.  
*it was stupid that you said*  
 ‘It was stupid that you said it/that.’
- b. Jag hörde att du sa \*(det).  
*I heard that you said*  
 ‘I heard that you said it/that.’

In (3a) the initial *det* is connected to a gap following the transitive verb *sa* (‘said’), inside the extraposed clause. (3b) shows that a gap in this position is ungrammatical if it is not linked to a filler. But where is the expletive *det<sub>x</sub>* which we have seen is obligatory in the postverbal position? In (4a) we have a version of this example with the object in situ. (4b) shows the expected version when the object is preposed; note *det<sub>x</sub>* after the first verb. (4b) is a grammatical sentence, but it is becoming less common. Instead speakers produce the version in (4c) without the expletive subject, marked here as  $\emptyset_x$ .

- (4) a. Det<sub>x</sub> var dumt att du sa det.  
*it was stupid that you said it*  
 ‘It was stupid that you said it/that.’
- b. Det<sub>i</sub> var det<sub>x</sub> dumt att du sa <sub>-i</sub>.  
*it was it stupid that you said*
- c. Det<sub>i</sub> var  $\emptyset_x$  dumt att du sa <sub>-i</sub>.

I began to notice such examples in informal conversations in the 1990’s. Since then it has become quite common and is now heard on the radio and occasionally found in print. Some additional examples are given in (5)–(6).

- (5) a. Det<sub>i</sub> är 'klart att du ska göra <sub>-i</sub>. (woman 50 years, 2001)  
*it is clear that you shall do*  
 ‘Of course you should do it.’
- b. Det<sub>i</sub> är ju 'roligt om hon får <sub>-i</sub>. (man 45 years, 2003)  
*it is PRT nice if she gets*  
 ‘It is nice if she gets it.’
- (6) Det<sub>i</sub> är ju så 'många som gör <sub>-i</sub>. (woman 50 years, 2005)  
*it is PRT so many that do*  
 ‘There are so many that do it.’

The examples have in common that there is an initial *det* followed by an impersonal or modal

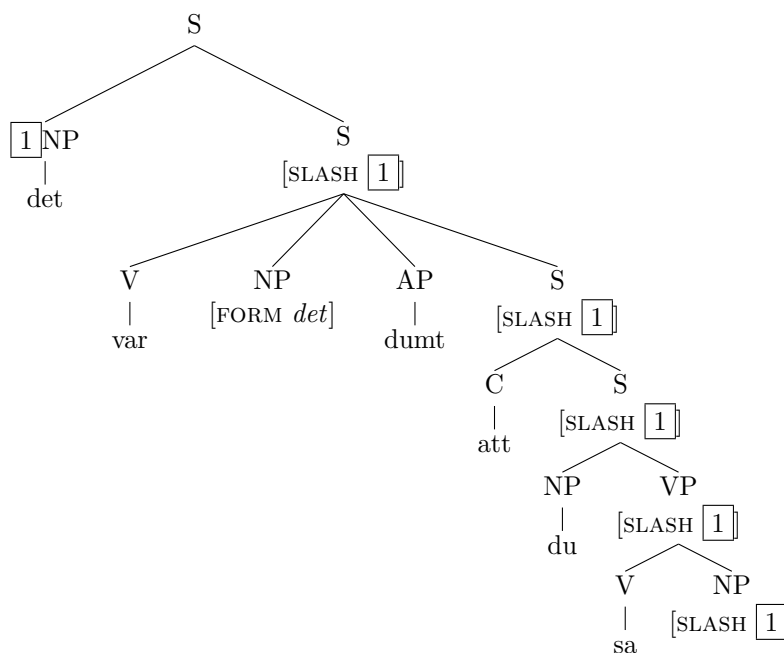
<sup>1</sup> In this respect, Swedish differs from German where so called ‘correlative *es*’ is only used in Spec,CP, (cf. Berman et al. 1998).

<sup>2</sup> See Engdahl (2012) for additional information about the authentic data and links to sound files.

verb or an evaluative or modal adjective which in turn is followed by a finite clause where there is a gap. The finite clause can either be a *that*-clause (4a, 5a), an *if*-clause (5b) or a relative clause as in the presentational construction in (6). The authentic examples are produced without hesitations or interruptions and do not cause any comprehension problems for the listeners.

Assuming that the initial *det* is the top of an extraction path which connects it to the missing object in the tensed clause, we can give an overview of the structure in (7), using a sort of hybrid notation.<sup>3</sup>

(7)



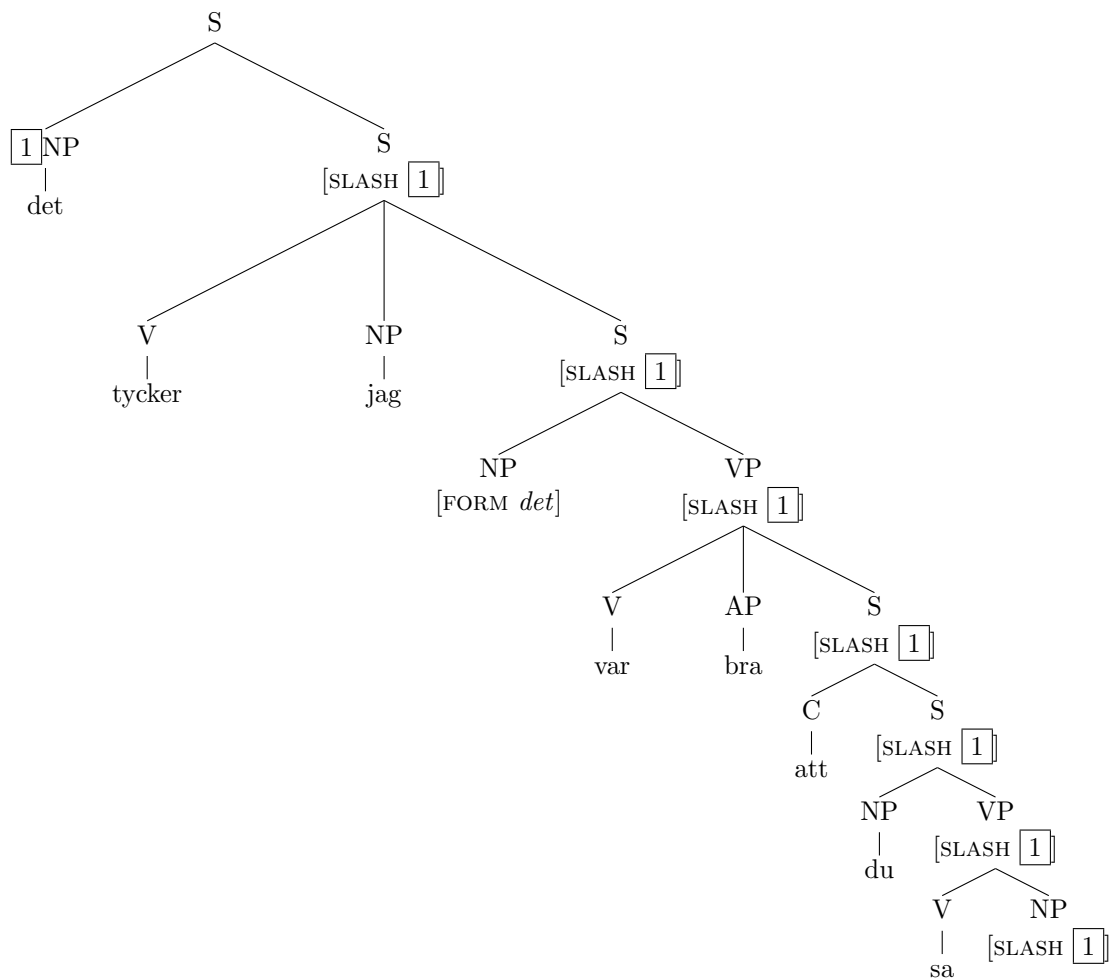
This overview tree encodes the information that there is a missing NP, whose PHON-value is *det*, linked to the missing object in the tensed clause via an extraction path indicated by SLASH. Crucially, there is another node which is constrained to have the form *det*, namely the NP node which licenses the extraposed clause. Whereas the grammar requires an overt *det* in both positions, it seems that the processor tends to overlook an unrealized expletive subject just in case it is dominated by a node with a SLASH feature whose PHON value is *det*.

This analysis makes the following prediction. We should find unrealized expletive subjects not just in the matrix clause but also further along on the extraction path. And this is exactly what we find in examples like (8). An overview structure is given in (9).

(8) Det<sub>i</sub> tycker jag Ø<sub>x</sub> var bra att du sa <sub>-i</sub>.  
*it think I was nice that you said*  
 'I think it was nice that you said it.'

<sup>3</sup> I don't see any reason for using a DOUBLE SLASH feature (see Müller 2008) to handle verb second in Swedish. Arguments for the flat structure in (7) can be found in Engdahl, Andréasson & Börjars (2004).

(9)



These kinds of examples are very interesting from the point of view of incremental processing. Let us look closer at how (8) might be processed.

- (10) a. Det<sub>i</sub> tycker jag <sub>-i</sub>  
*it think I*
- b. Det<sub>i</sub> tycker jag [<sub>S</sub>-<sub>i</sub> var bra  
*it think I was nice*
- c. Det<sub>i</sub> tycker jag [<sub>S</sub>Ø<sub>x</sub> var bra [<sub>S</sub> att du sa <sub>-i</sub>]]  
*it think I was nice that you said*

Assuming that the processor has recognized that the initial *det* is not a likely subject for the verb *tycker* ('think'), which requires an animate subject, the first assumption might be that it is the object, as indicated in (10a).<sup>4</sup> When another finite verb (*var*) is recognized, the

<sup>4</sup> In Swedish, *tycker* often takes *det* as an object:

- (i) Jag tycker också det.  
*I think also it*  
'I also think so.'

processor needs to abandon this assumption and instead postulate a clause and a missing subject (10b). But when the processor recognizes that there is an object gap after *sa* in the most embedded clause, the filler must be linked to this gap. In addition, the empty subject in the intermediate clause has to be reinterpreted from a gap to an empty expletive, as shown in (10c).

The fact that examples like (8) are both produced and understood without any noticeable effort shows that the processor must be able to reanalyze the gaps very quickly. The utterances are produced rapidly, without hesitations or self-corrections, and there is no indication from the way the listeners' react that they have trouble understanding.

Another challenge for the processor is shown in (11). This is from an interview on Swedish radio where a former heroin addict is talking about quitting. (.) indicates short pause

(11)

B: de e Ø ju alltid nån nån annan som vill e<sub>1</sub> (.)  
*it is PRT always some some else that wants*  
 att man ska göra e<sub>2</sub> (.)  
*that one shall do*

When the processor recognizes the verb *vill*, which is followed by a short pause, it presumably connects the gap with the filler *det*. But when the speaker resumes and produces yet another subordinate clause with a gap after *göra*, the processor must reanalyze the *att*-clause as the complement of *vill* and the gap after *göra* as the bottom of the extraction path.

These examples all start with *det*. Since *det* can be either expletive or referential, almost all examples remain ambiguous until most of the utterance is processed. This is true as well for *tough* constructions in English, as in (12), where the final status of *it* as referential (*it<sub>i</sub>*) or expletive (*it<sub>x</sub>*) is shown after the arrow.

- (12) a. It is hard. → *it<sub>i</sub>*  
 b. It is hard to live. → *it<sub>x</sub>*  
 c. It is hard to tell   . → *it<sub>i</sub>*  
 d. It is hard to tell the truth. → *it<sub>x</sub>*

The same applies to *det* in Swedish *tough* constructions. In addition we get the minimal pairs in (13) which I believe can be seen as an extension of the *tough* construction (see Engdahl 2012).

- (13) a. Det var dumt att du kom. → *det<sub>x</sub>*  
*it was stupid that you came*  
 b. Det var dumt att du sa   . → *det<sub>i</sub>*  
*it was stupid that you said*  
 c. Det är ju så många som gör det. → *det<sub>x</sub>*  
*it is PRT so many that do it*

---

(ii) Det tycker jag också.  
*it think I also*

- d. Det är ju så många som gör \_ . → det<sub>i</sub>  
*it is PRT so many that do*

Given that the processor can't tell whether an initial *it* or *det* is referential or expletive until the end of the utterance, it must presumably keep both options open. If we go back to the structure in (7), this would mean that the processor recognizes that there is an empty subject gap, suitable for *det*, immediately after the first verb, but that it also retains *det* as a potential filler for a possible gap further along. It is interesting that keeping both options open does not seem to be problematic. I'm not aware that listeners – or speakers, if we include the planning perspective – experience any problems with examples like (12) and (13).

Why then are the examples in (13) not used in English? After all, the English expletive *it* is identical to the referential *it*, which is a precondition for this construction. And English displays a similar ambiguity in *tough*-constructions as shown in (12). I believe that this has to do with the way preposing is used in Swedish, in particular with the way pronouns are preposed. In Swedish, unaccented pronouns are very often fronted, as a way of establishing coherence, especially in conversations, but also in texts. Consider some ways one can answer the question in (14).

(14) Do you like to make bread?

- a. Ja, det är trevligt. / Yes, it is nice.  
 b. Ja, det tycker jag \_ är trevligt.  
*yes it think I is nice*  
 c. \* Yes, it I think \_ is nice.  
 d. # Yes, that I think \_ is nice.  
 e. Ja, det tycker jag Ø är trevligt att jag får göra \_.  
*yes it think I is nice that I get do*  
 'Yes, I think it is nice that I get to do it.'

The initial subject *det* and *it* in (14a) refers to the activity introduced in the question *to make bread* and is normally unaccented. (14b) with *det* linked to the subject position of an embedded clause is another common way of answering such questions in Swedish. In English, this is not possible (14c). Something like (14d) with an accented *that* would be possible, but it is not felicitous in this context. Finally (14e) shows how the reply is easily extended to the type of example I am focussing on in this talk with a missing expletive *det* on the extraction path.

Preposing of non-contrastive pronouns is found in all the Scandinavian languages, but interestingly, only Norwegian allows expletives to be dropped as in Swedish. This suggests that there is a certain amount of conventionalization involved. In Danish it is not possible to leave out *det* in matrix clauses, but intermediate expletives as in (8) may be unrealized.

German is interesting in this context. The weak object pronoun *es* can not normally be fronted but Theiler & Bouma (2012) have recently shown that it is possible, just in case the initial *es* can also be interpreted as an expletive, as it would be in presentational constructions when the subject is non-topical.



- (15) Wir könnten dir helfen aber: ich denke, es wird keiner machen.  
*we could you help but I think it will nobody do*  
 ‘We could help you, but I think nobody will.’ (Theiler & Bouma 2012:(18b))

Theiler & Bouma analyse the initial *es* as an amalgamation of an object *es* and a presentational expletive. Preposing of *es* from a subordinate clause, as in the Swedish examples discussed here, is impossible.

- (16) Es war blöd [<sub>CP</sub> daß du \*(es) gesagt hast]  
*it was stupid that you it said has*

To round off: I have described a new and spreading construction in Swedish, where expletive *det* is optional just in case there is an extraction path, headed by a referential *det*, spanning it. It is of course not an accident that this happens precisely when the filler is *det*, i.e. indistinguishable from the expletive *det* that licenses the extraposed clause in which the ultimate gap is found. Although I have mentioned processing strategies several times, there is actually very little known about how listeners, and speakers, handle such dependencies. There is clearly a need for experimental studies, preferably on-line studies using materials that come as close as possible to the authentic materials discussed here. This might give us some insights into the processing load at various points on the extraction path.

### References

- Berman, J., S.Dipper, Ch.Fortmann & J.Kuhn (1998) Argument clauses and correlative *es* in German – deriving discourse properties in a unification analysis. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of LFG98*. Stanford: CSLI Publications.
- Engdahl, Elisabet (1985) Parasitic gaps, resumptive pronouns, and subject extractions. *Linguistics* 23:3–44
- Engdahl, Elisabet (2012) Optional expletive subjects in Swedish. *Nordic Journal of Linguistics* 35:99–144.
- Engdahl, Elisabet, Maia Andréasson & Kersti Börjars (2004) Word order in the Swedish midfield – an OT approach. In F.Karlsson (ed.) *Papers from the 20th Scandinavian Conference of Linguistics*. University of Helsinki, Department of General Linguistics, Publications No. 36, 1–13. <<http://www.ling.helsinki.fi/kielitiede/20scl/proceedings.shtml>>
- Gazdar, Gerald (1981) Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12:155–84.
- Kayne, R.S. (1984) *Connectedness and Binary Branching*. Dordrecht: Foris.
- Müller, Stefan (2008). *Head-Driven Phrase Structure Grammar. Eine Einführung*. Tübingen: Stauffenburg Verlag.
- Theiler, Nadine & Gerlof Bouma (2012) Two for the Price of One: An LFG treatment of sentence initial object *es* in German. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of LFG12*. Stanford: CSLI Publications. <<http://csli-publications.stanford.edu/LFG/>>
- Zaenen, Annie (1983) On Syntactic Binding. *Linguistic Inquiry* 14:469–504.



**GU-ISS**, Forskningsrapporter från Institutionen för svenska språket, är en oregelbundet utkommande serie, som i enkel form möjliggör spridning av institutionens skriftliga produktion. Det främsta syftet med serien är att fungera som en kanal för preliminära texter som kan bearbetas vidare för en slutgiltig publicering. Varje enskild författare ansvarar för sitt bidrag.

**GU-ISS**, Research reports from the Department of Swedish, is an irregular report series intended as a rapid preliminary publication forum for research results which may later be published in fuller form elsewhere. The sole responsibility for the content and form of each text rests with its author.

**Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet**  
**Research Reports from the Department of Swedish**

**ISSN 1401-5919**

[www.svenska.gu.se/publikationer/GU-ISS](http://www.svenska.gu.se/publikationer/GU-ISS)