Acceptability and processing of long-distance dependencies in Danish

Poulsen, Mads

doi:10.1017/S0332586508001832

Acceptability and processing of long-distance dependencies in Danish

Mads Poulsen

Long-distance dependencies have been the object of much theoretical interest in the Scandinavian languages and in general, but the empirical foundation for theorizing has been limited. The present paper investigates extraction from complement and adverbial clauses in Danish using acceptability judgment and reading-time measures. Extraction from adverbial clauses was found to be rated near the bottom of the scale and to be associated with a processing cost. This was also true of extraction in adverbial clauses in semantically cohesive sentences, which Jensen (2001a, b) has suggested is acceptable. It is concluded that under the conditions investigated extraction from adverbial clauses in Danish is associated with a processing cost and very low acceptability ratings, despite semantic cohesion.

Keywords acceptability, Danish, extraction, grammaticality, island constraints, judgment, language processing, long-distance dependencies

Mads Poulsen, Department of Nordic Studies and Linguistics, University of Copenhagen, Njalsgade 120, DK-2300 Copenhagen, Denmark
m.poulsen@hum.ku.dk

1. INTRODUCTION

This paper is an empirical investigation of the acceptability and processing of certain long-distance dependency constructions in Danish. A long-distance dependency is here defined as a grammatical relation holding between a predicate and an argument that is not positioned in the same clause as the predicate. Alternatively it can be defined as extraction across a clause boundary. I shall use both terminologies interchangeably.

English allows long-distance dependencies, for example, through *wh*-question formation, as in (1), and in Danish such dependencies can also be established through fronting of a non-*wh*-constituent, as in (2).

(1) What $_i$ do you think [John bought $\_i$].
(2) Frakken$_i$ tror   jeg [John købte $\_i$].
     coat.DEF$_i$ think I   John bought $\_i$

In both these examples a dependency exists between the subordinate clause verb *bought/købte* and the first element of the matrix clause.

There are, however, restrictions on the setting up of such long-distance dependencies in that dependencies cannot be formed out of certain clauses, termed

ISLANDS since Ross (1967). It has been suggested that there exist a number of universal, syntactic constraints on when long-distance dependencies can be formed. One such constraint is the Complex NP constraint, prohibiting dependency formation into relative clauses dominated by an NP with a lexical head noun:

(3) *What$_i$ do you know [the man [who bought _$_i$]]?

Other islands include sentential subjects and adjunct clauses. These constraints have been generalized under the notion of the subjacency condition (Chomsky 1973).

In the 1970s and the 1980s, the Scandinavian languages sparked a great deal of interest in this context because they appeared to contradict the proposed universal constraints. Specifically, long-distance dependencies into relative clauses were reported to be possible in Danish (Erteschik-Shir 1973, 1982; Erteschik-Shir & Lappin 1979; Jakobsen 1995; Jensen 2001a), Swedish (Allwood 1982; Anderson 1982; Engdahl 1982, 1997) and Norwegian (Taraldsen 1982).[1] An example of this from Swedish is provided in (4).

(4) Den teorin$_i$ känner jeg [ingan [som tror   på _$_i$]].[2]
     that theory$_i$ know I   no.one who believe in _$_i$

It has furthermore been argued that the conditions on long-distance dependencies in the Scandinavian languages and in general are pragmatic, not syntactic, in nature (e.g. Erteschik-Shir & Lappin 1979, Deane 1991, Van Valin 1995, Engdahl 1997), because acceptability varies with sentence context and non-syntactic properties of the sentences, e.g. the matrix verb:

(5) Det$_i$ har   jeg mødt [mange [der har   gjort _$_i$]].
     that$_i$ have I   met   many   who have done _$_i$
(6) *Det$_i$ har   jeg drillet [mange [der har   gjort _$_i$]].
      that$_i$ have I   teased   many   who have done _$_i$ (Erteschik-Shir 1982)

1.1 Methodological issues

Although there has been a fair amount of theorizing as to which constraints hold and as to their nature, there has been relatively little systematic work on the empirical basis of the proposed theories in the linguistics literature. Most of the theoretical work is based on what we must assume is acceptability judgments made by the author. This is sometimes supplemented with real-life observations and consultations with native speakers, but details on the nature of the data are usually limited. Judgments made by linguist authors is probably the most common data source in language description and linguistic theorizing, but there has been a rising concern on the validity of the method (Labov 1975, Schütze 1996, Cowart 1997, Poulsen 2005). Among the concerns are suggestions that linguists might differ from non-linguists in their grammar or be influenced by theoretical preferences (Dąbrowska 1997, Snow & Meijer 1977), or

that sensitivities are dulled through repeated exposure (Snyder 2000). Linguistics would grind to a halt if linguists' judgments were banned, but in grey-area cases and cases of controversy it seems reasonable that data are supplied with some measure and specification of reliability beyond the upstanding character of the linguist (cf. Schütze 1996, Poulsen 2005, Phillips & Wagers 2007 for further discussion).

Over the last 20 years a number of psycholinguistic studies have investigated the processing of island constraints, primarily in English, using various experimental methods (cf. Poulsen 2005, Phillips 2006 for reviews). This research has supported the existence of certain constraints, e.g. on extraction out of adverbial clauses (McKinnon & Osterhout 1996), interrogative clauses (Kluender & Kutas 1993) and relative clauses (Traxler & Pickering 1996), and with regard to parasitic gaps (Phillips 2006). But the focus of these studies has been to investigate WHEN constraints influence processing in on-line comprehension. The general picture today is that island constraints influence processing quickly during the comprehension of a sentence, e.g. to rule out dependency formation into relative clauses (Traxler & Pickering 1996) or to trigger event-related potential responses (ERPs) at the first sign of a constraint violation (McKinnon & Osterhout 1996).

In the area of Scandinavian extraction, Allwood (1982) provides more detail on data than most by reporting that he checked the acceptability of all sentences in his paper with a group of 19 native speakers, and for all sentences at least one speaker agreed with his judgment. This, however, leaves the possibility that all the judgments of the author were rejected by 95% of the speakers consulted. That was probably not the case, but the point is that the level of methodological detail leaves us with very little information on the extent of the empirical basis of a generalization. Such a situation is especially problematic for the non-native linguist readers who cannot consult their own intuition.

There are some studies of Scandinavian long-distance dependencies with a high level of empirical detail. Jensen conducted a corpus study of long-distance dependencies in Danish (Jensen 1998, 2001a, b). She searched through 18 hours of a corpus of spoken Danish and found 230 examples of long-distance dependencies. Of these 230 extractions, 219 were from complement clauses while ten were from relative clauses; one example was from an adverbial clause. In general, these findings support the widely-held assumption that extraction from complement clauses is acceptable. But it also illustrates the limits of this type of corpus study, namely that it is difficult to interpret the status of structures that occur infrequently; should we conclude based on one occurrence that extraction out of adverbial clauses is a possibility in Danish, or should the example be considered the result of a production error? Authentic examples are generally highly regarded as a data type, because they reflect real-world language use. But considering that we know that people sometimes make production errors, e.g. by switching the order of words (cf. e.g. Bock & Levelt 1994 for an overview), it is difficult to determine on principled grounds whether very rare examples of a

certain construction reflect planning errors or the grammar of a language. The same problem holds for non-occurring structures. How do we know whether the lack of occurrence of a structure is due to ungrammaticality or, say, bad sampling?

Jensen's ten examples of relative clause extractions should presumably not be considered production errors. But ten examples provide a limited basis for generalizing about the conditions on relative clause extraction. Jensen acknowledges this by using additional material to augment her analyses, primarily her own judgments.

Creider (1987) exemplifies a different approach. He did a questionnaire-based acceptability judgment study with 59 Norwegian native speakers, and manipulated a large number of different factors that could possibly affect the acceptability of extraction, such as the type of clause extracted from. I will refer to this methodology as objective acceptability judgments (cf. Cowart 1997). The methodology has the advantage over corpus studies that specific variables can be controlled tightly, low-frequency phenomena can be investigated, the problem of non-occurrence can be sidestepped, and the inclusion of many raters guaranties intersubjectivity of the results (cf. Bard, Robertson & Sorace 1996, Schütze 1996, Cowart 1997, Poulsen 2005 for further methodological discussion).

The study does, however, suffer from a few methodological problems. It was very ambitious in the sense that it appears to include 44 different structural experimental conditions. But each condition was only instantiated by one sentence. A consequence of this is that although there was a high number of informants, which ensures good generalizability to the population of speakers, we cannot in the same way generalize from one stimulus sentence to the population of sentences, as it were, with the same structural features. Furthermore, many conditions were not manipulated as minimal pairs, and it appears that the order of presentation was not randomized – all participants judged the sentences in the same order, and the order of sentences was related to their condition. For example, all extractions from adverbial clauses seem to have been placed last in the questionnaire. This makes it difficult to interpret the result of the study that extraction from adverbial clauses were rated surprisingly acceptable, because the high rating could reflect the fact that the participants already had rated a lot of similar extraction sentences, and that their intuitions were somewhat dulled by this. Similarly, although Creider did examine whether context improved acceptability ratings, and found that they did, the result was confounded by the fact that sentences were first rated without context, and then after all sentences had been rated, they were rated again with context. Again it is impossible to determine whether the effect of context was really an effect of familiarity, especially since there are few details as to how the contexts provided were supposed to improve the acceptability.

The purpose of the present paper is to provide robust data within a limited area of the extraction phenomenon in Danish using both off-line and on-line experimental measures. The idea is that the results can provide a baseline for further investigations.

1.2 The theoretical issue

Although the data on long-distance dependencies in Scandinavian seem sufficient to document e.g. the existence of relative clause extraction, and to give some ideas on the constraints on extraction, it is not clear that they are sufficient to actually test more sophisticated theories of constraints on extraction. Even fairly basic questions are difficult to answer on empirical grounds. A good example is adverbial clause extraction in Danish. Togeby (2003) states that such extractions are unacceptable. And as noted above, they are almost, but not quite, absent in Jensen's corpus. But Jensen does report (2001a:25) having observed a few outside the corpus, e.g.:

(7) det$_i$ blev   hun smaddersur fordi   jeg sagde _$_i$
    that$_i$ became she very.angry because I   said _$_i$

Similar examples have informally been noticed by other linguists, including myself. On the other hand, there are some extractions out of adverbial clauses that Jensen consulting her own judgment considers unacceptable, e.g.:

(8) *Det vindue$_i$ går jeg ud hvis du åbner _$_i$.
     that window$_i$ go I   out if   you open _$_i$

To account for this, Jensen proposes the generalization that extraction from adverbial clauses is acceptable if there is a high degree of semantic cohesion between the two clauses. Following Van Valin & LaPolla (1997), she defines semantic cohesion as 'the extent to which a given construction expresses facets of a single event, action or state of affairs or discrete events, actions or state of affairs' (Van Valin & LaPolla 1997:478 quoted in Jensen 2001a:26). The actual data the generalization is based on are, however, very limited. They consist of a few authentic examples (which were not from a systematic corpus) and her own subjective judgments. Jensen's generalization does not stand out because of its theoretical clarity, but it is worth investigating because it is an attempt to account for anecdotal observations of extraction out of adverbial clauses made by Jensen, myself and other linguists. On a personal note, the reason why I have actually noticed some of these examples in running discourse has been that I had difficulty comprehending them and subsequently found them odd. Now, after having worked on the subject for a few years, I do not find them odd anymore. So whom should we trust?

When the empirical status of the fairly basic question of the acceptability of extraction from adverbial clauses is somewhat shaky, the foundation of rather complicated proposals on, for example, the role of context and information structure status, becomes even harder to evaluate.

The aim of the rest of paper is to illustrate a methodological stringent line of empirical work that can be pursued in the investigation of conditions on extraction. Experiments 1a and 1b provide data on the basic distinction between extraction out of adverbial clauses versus extraction out of complement clauses. The second set

of experiments, 2a and 2b, investigates Jensen's proposal that extraction out of an adverbial clause with a high degree of semantic cohesion with its matrix clause is acceptable.

2. EXPERIMENTS 1A AND 1B

2.1 Experiment 1a: acceptability judgments

The first experiment was an objective acceptability judgment experiment. The purpose of the experiment was to establish whether adverbial clause extraction sentences were rated lower on acceptability than the corresponding complement clause extraction sentences.

2.1.1 Design

The experiment had a 2 by 2 factorial design, crossing type of construction (extraction vs. non-extraction) with type of subordinate clause (complement vs. adverbial). The type-of-clause factor was necessarily confounded with the type of matrix verb such that in the complement clause conditions, the matrix verbs were of a type that only takes complement clauses as the second argument (comp-verbs), while in the adverbial clause conditions, the matrix verbs were intransitive. This design yielded four experimental conditions illustrated below:

Extraction, complement clause

(9) Den mappe   tror jeg at   de glemmer på kontoret.
     that briefcase think I   that they forget   at office.DEF

Extraction, adverbial clause

(10) Den mappe   besvimer jeg hvis de   glemmer på kontoret.
      that briefcase faint   I   if   they forget   at office.DEF

Non-extraction, complement clause

(11) Jeg tror at   de glemmer den mappe   på kontoret.
      I   think that they forget   that briefcase at office.DEF

Non-extraction, adverbial clause

(12) Jeg besvimer hvis de   glemmer den mappe   på kontoret.
      I   faint   if   they forget   that briefcase at office.DEF

The conditions of primary interest were the extraction conditions. The non-extraction conditions were included to control for the possibility that the two

extraction conditions might differ in overall semantic plausibility because the same general sentence template had to accommodate two different matrix verbs and clausal relationships. It would thus be possible that any difference in acceptability score would be a consequence of differences in semantic plausibility rather than differences associated with extraction constraint violation. The non-extraction conditions control for this because they are truth-conditionally equivalent to the corresponding extraction conditions. The importance of closely matching the sentence form between the conditions will become clear in the context of Experiment 1b.

2.1.2 Materials

Sixteen sentence items were constructed. Each item consisted of four different sentences corresponding to the four experimental conditions. The sentences in (9)–(12) above illustrate the four sentences of one of the items. The rest of the items are provided in Appendix A. All sentences followed the same general form exemplified above. In the extraction conditions, the specification of the sentences was as follows:

> fronted NP (demonstrative pronoun + noun) + present tense matrix verb + first-person, nominative pronoun + complement or adverbial subordinator + nominative pronoun + transitive, present tense verb + sentence-final PP

The fronted NP functioned as the direct object of the subordinate clause verb. Demonstratives were chosen because they intuitively seem more appropriate in these extraction sentences than e.g. indefinites, as has also been observed for Swedish (Anderson 1982). In the adverbial clause conditions, *hvis* 'if' was used as the subordinator, while *at* 'that' was used in the complement clause conditions. The non-extraction sentences had the same form with the following two exceptions: the subordinate direct object appeared in its canonical position after the subordinate verb, and the matrix subject appeared before the matrix verb.[3]

To create the 16 items, eight verbs taking clauses as complement (comp-verbs) and eight intransitive verbs were used to form eight arbitrary pairings of an intransitive verb and a comp-verb. Each pair was used in two items. The sentences into which the verbs were inserted were constructed to be as semantically plausible as possible. This was difficult because each sentence template had to accommodate both a comp-verb and an intransitive verb, and the balance may not always have been completely successful. But the non-extraction conditions were, as noted above, included to control for semantic implausibility in the data analysis.

The sentences were distributed across four counterbalanced lists: each list contained one sentence from each item with four sentences from each condition. The same verb occurred only once in each list. Every list additionally had 32 unrelated filler sentences. These represented an even spread of acceptability levels. The acceptability levels of these sentences were estimated on the basis of experiences

from an earlier pilot study. Each list thus contained 48 sentences. Every list had four different pseudo-randomized orders of presentation; there were always two fillers between each experimental item, but the order of fillers and experimental items was randomized. The result was 16 different questionnaires. Every informant filled out one questionnaire, ensuring that no participant saw the same main verb more than once and that each participant saw precisely one sentence from each item and an equal number of sentences from each condition. Each questionnaire was filled out by at least two participants.

All of these measures were taken to eliminate the risk that order of presentation or familiarization with individual sentences would confound the results.

2.1.3 Procedure and participants

The experiment was a questionnaire-based acceptability judgment experiment (cf. Cowart 1997, Poulsen 2005). Thirty-six second-year high school students from Ingrid Jespersens Gymnasium in Copenhagen participated on a voluntary basis. Data from only 32 of the informants were used (see the results section for details). The mean age of this subset was 17;5, ranging from 17 years to 18 years.

The questionnaires presented the sentences separately. Below every sentence was a sequence of numbers from 1 to 5. Informants were instructed to rate each sentence on a scale from 1 to 5, 5 being completely natural (Danish: *naturlig*), 1 completely unnatural (Danish: *unaturlig*), and to mark their rating by circling the number below the sentence corresponding to their judgment. The context story for the task was that participants should imagine themselves as teachers of Danish to immigrants. This was done to provide a plausible context for what could otherwise be considered an odd task. They were additionally instructed to consider the sentences one at a time, to give every sentence a rating on the scale, and not to change a rating once it was marked.

2.1.4 Predictions

The predictions are fairly simple. If extraction out of adverbial subordinate clauses is problematic, we expect adverbial extraction conditions to be rated less acceptable than the complement extraction conditions. We would also expect an interaction between subordinate clause type and extraction such that the difference between the extraction adverbial condition and the non-extraction adverbial condition would be greater than the difference between the two complement conditions. Such an interaction would show that the extraction effect is not due to differences in plausibility between the truth-conditional semantics of the adverbial and complement conditions.
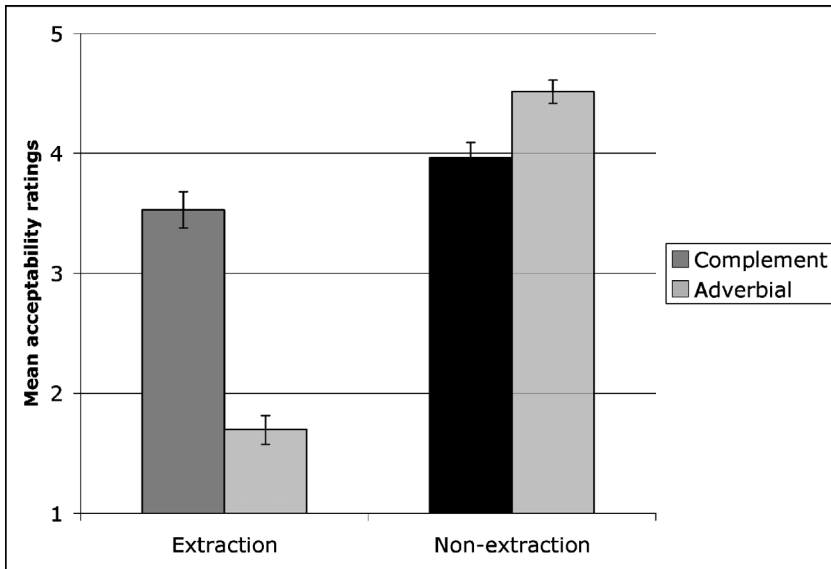
Figure 1. Mean acceptability ratings in Experiment 1a.

2.1.5 Results and discussion

The data from two participants were discarded because they listed languages other than Danish as their mother tongue. Two additional participants were pulled out semi-randomly to even the number of participants in each experimental list.

The average ratings for the four conditions are displayed in Figure 1.

An analysis of variance with construction type and matrix verb as factors showed main effects of extraction ($F_1(1,31) = 548.85$, $p < 0.001$; $F_2(1,15) = 33.87$, $p < 0.001$) and subordinate clause type ($F_1(1,31) = 45.57$, $p < 0.001$; $F_2(1,15) = 24.80$, $p < 0.001$) and an interaction between the two factors ($F_1(1,31) = 114.65$, $p < 0.001$; $F_2(1,15) = 141.19$, $p < 0.001$). The main effect of extraction suggests that extraction sentences were rated lower in both clause type conditions. One-way analyses of variance also showed that extraction sentences were rated lower both in adverbial clause conditions ($F_1(1,31) = 417.39$, $p < 0.001$; $F_2(1,15) = 441.01$, $p < 0.001$) and in complement clause conditions ($F_1(1,31) = 12.27$, $p < 0.001$; $F_2(1,15) = 12.52$, $p < 0.01$).

The results were, in other words, largely as expected. Extractions out of adverbial clauses were rated significantly lower than extraction out of complement clauses. The interaction between the two factors means that this effect cannot be explained as an effect of the truth-conditional semantics of the adverbial extraction sentences being less plausible than their complement extraction counterparts. In fact, it seems that
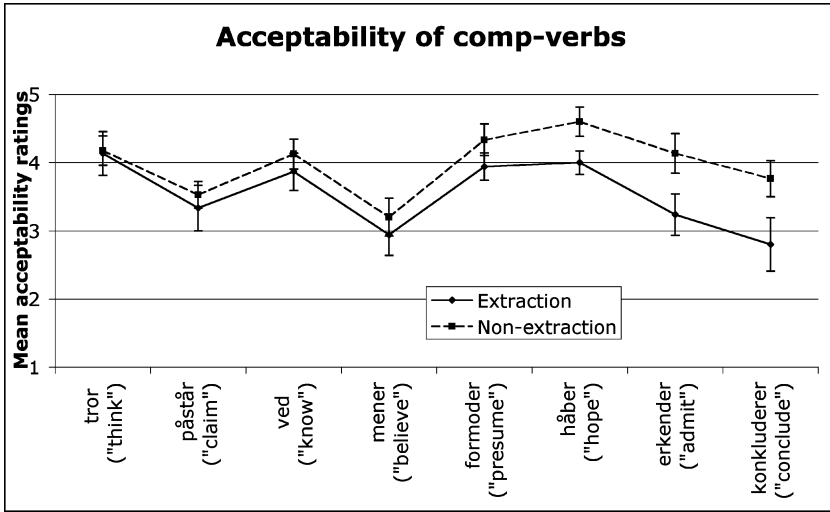
Figure 2. Acceptability of individual comp-verbs.

without extraction, the adverbial subordinate clause sentences were rated higher than the complement clause sentences. This is in all probability due to the constraint that the pairs of comp-verbs and intransitive verbs had to fit into the same sentence frames. Even though I attempted to select verbs that fit equally well in the sentence frames, this constraint was particularly difficult to satisfy.

One thing worth noting is that extraction also affected the acceptability in the complement clause conditions. This raises the question of whether or not such extractions should also be considered problematic. Although the effect is statistically significant, it is relatively small, and the mean score for the extraction condition is still in the top half of the scale. But the extraction sentences could be considered less acceptable than the non-extraction counterparts. Since the analyses were averaged over eight different comp-verbs, an explanation of some of the effect could be that the verbs differ with regard to the extent to which they allow extraction. Figure 2 presents the averages of the individual comp-verbs in the extraction and non-extraction conditions.

The verbs are presented from left to right in the order of the degree to which they allow extraction. Degree of extractability was computed by subtracting the average acceptability of the verb in the extraction condition from the average acceptability of the verb in the non-extraction condition.[4] The verb *tror* 'think, believe' is the verb that best allows extraction, whereas *konkluderer* 'conclude, infer' is the one where the difference between the extraction and the non-extraction conditions is the largest. The differences range from 0.04 to 0.96, with lower numbers signifying higher degree
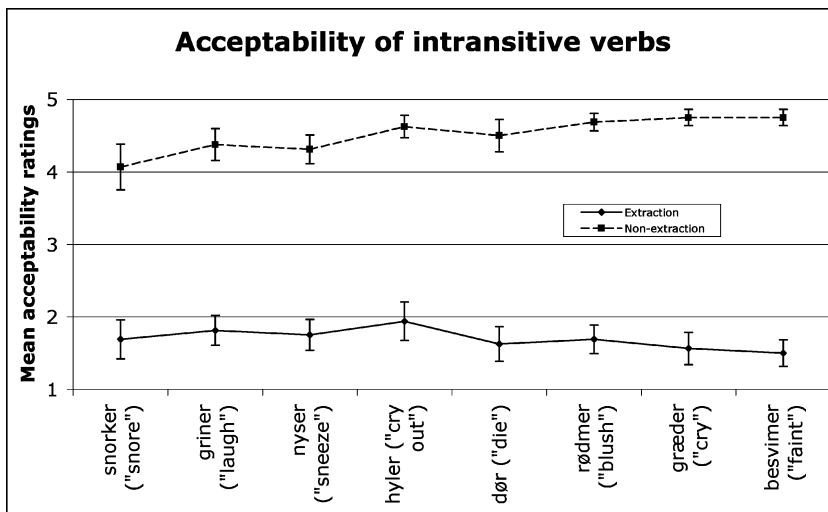
Figure 3. **Acceptability of individual intransitive verbs.**

of extractability. To get a context for understanding the size of the differences, the same type of analysis was done on the intransitive verbs, as shown in Figure 3.

The differences between the extraction conditions and the non-extraction conditions range from 2.38 to 3.25. Thus, even though we see some variability in extractability between the comp-verbs, these verbs all have much higher extractability than the intransitive verbs. The experiment was not designed to investigate differences between the individual verbs, so no clear conclusions based on these analyses can be drawn. But it does appear that certain verbs account for a good portion of the negative extraction effect for the comp-verbs.

2.2 Experiment 1b: self-paced reading

One problem with acceptability judgments is that it may not be entirely clear to the test subjects what they are really supposed to do, which may for example make them prone to irrelevant normative considerations.

To supplement the acceptability judgments for the sentences in Experiment 1a, I conducted a reading-time experiment with the same sentences. The assumption was that unacceptable sentences would cause interpretation difficulties which would translate into prolonged reading times. The basic task for the participant was to read sentences one word at a time with the purpose of understanding the sentence well enough to be able to answer a comprehension question for the sentence. This methodology provides a reaction time measure for each word in a sentence. Elevated reaction times on specific words can be detected by comparing the reaction times on

specific word positions in minimal pair sentences. The method has been widely used in studies of sentence processing to detect processing difficulty, and I have elsewhere (Poulsen 2005) discussed in more detail the foundations for using the methodology in linguistic investigations of acceptability.

Although the method of reading is relatively unnatural, informants get used to it fairly quickly, and the actual task is easy to explain and understand, unlike the judgment task. In this sense the task is relatively natural.

2.2.1 Participants and procedure

Thirty-two students from various departments at the Faculty of Humanities, University of Copenhagen, participated on a voluntary basis. All were recruited from undergraduate-level courses. All reported Danish as their first language. Mean age was 25;3, ranging from 19 years to 42 years.

The experiment was conducted on an Apple Macintosh computer with an attached Carnegie Mellon button box. The PsyScope software package (Cohen et al. 1993) was used to present stimuli and record reaction-time latencies between the presentation of a word and the button press for the next word. Sentences were presented one word at a time in a self-paced moving window paradigm (Just, Carpenter & Woolley 1982). After each sentence, participants had to answer a comprehension question. Participants were instructed to read at a fairly quick pace, but with the purpose of understanding the sentence so that they could answer the question. The function of the comprehension questions was to provide a sensible task for the participants and to ensure that they read with the purpose of understanding the meaning of the sentences.

2.2.2 Design and materials

The 2 (construction: extraction vs. non-extraction) by 2 (clause type: complement vs. adverbial) factorial design and the stimuli from Experiment 1a were used.

2.2.3 Predictions

The general prediction is that the extractions out of adverbial clauses should be read more slowly than complement extractions, but the question is when the difference should show up. In an ERP study using similar sentences in English, McKinnon & Osterhout (1996) found evidence of processing difficulties at the subordinator. On the other hand, the authors of a comparable study (Kluender & Kutas 1993, Kluender 1998) concluded that the extraction-related breakdown was located at the gap-site, when the extracted element had to be integrated into the meaning representation. We would thus expect difficulties to show up at one of these positions.
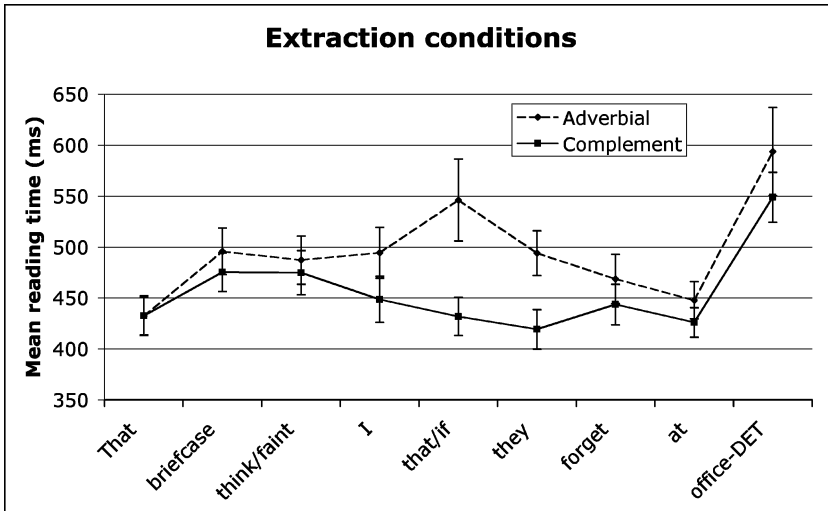
Figure 4. Experiment 1b mean reading times in extraction conditions.

2.2.4 Results and discussion

For each word position, data points above or below three standard deviations from the condition mean were replaced with the cut-off value to reduce variance from outliers. This affected 1.77 percent of the data.

Figure 4 summarizes the mean reading-times data for the extraction conditions, while Figure 5. summarizes the mean reading times for the non-extraction conditions. The data are presented in two figures because word-for-word comparisons are not straightforwardly possible across the construction factor. An informal inspection of the figures suggests that reading times in the extraction conditions diverge from each other around the subordinator, while there does not appear to be any difference in the non-extraction conditions. This impression was verified by a 2 by 2 analysis of variance of reading times at the subordinator. There were main effects of both verb type ($F_1(1,31) = 12.89$, $p < 0.001$; $F_2(1,15) = 6.20$, $p < 0.05$) and construction ($F_1(1,31) = 6.22$, $p < 0.05$; $F_2(1,15) = 6.75$, $p < 0.05$), and interaction ($F_1(1,31) = 11.45$, $p < 0.01$; $F_2(1,15) = 5.37$, $p < 0.05$). The interaction at the subordinators is displayed in Figure 6.

As can be observed, the extraction factor has little impact on mean reading times at the subordinator in the complement conditions. The main effect of construction seems to be carried primarily by the adverbial extraction condition, and a one-way analysis of variance on the complement verbs revealed no effect (both Fs less than 1).
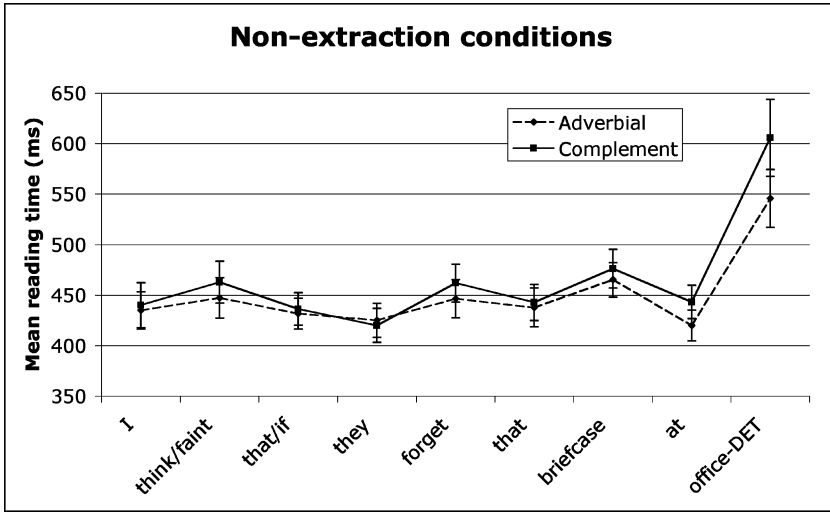
Figure 5. Experiment 1b mean reading times in non-extraction conditions.
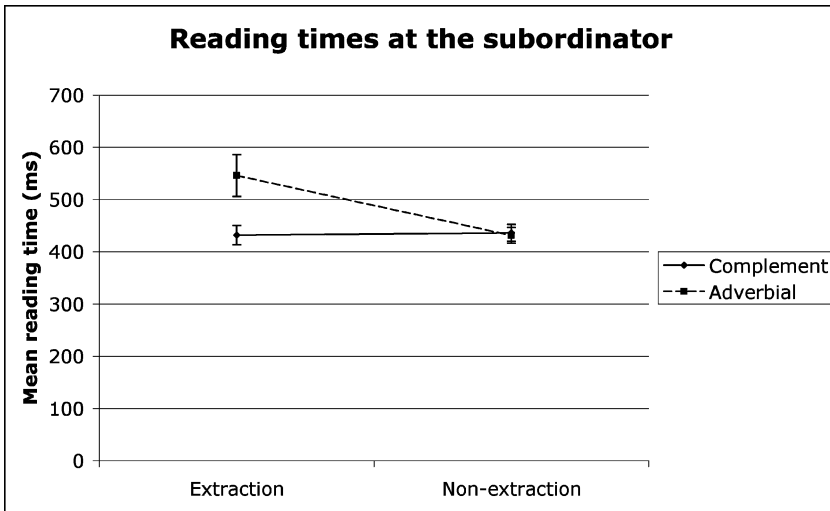

Figure 6. Experiment 1b mean reading times at the subordinator.

The same pattern was observed at the word following the subordinator, the subordinate subject: there was a main effect of verb type ($F_1(1,31) = 22.36$, $p < 0.001$; $F_2(1,15) = 7.40$, $p < 0.05$) and construction ($F_1(1,31) = 9.24$, $p < 0.01$; $F_2(1,15) = 4.12$, $p = 0.06$), and the interaction was significant by subjects and

marginally significant by items ($F_1(1,31) = 9.00$, $p < 0.01$; $F_2(1,15) = 3.79$, $p = 0.07$). At the subordinate verb there were no significant main effects or interactions (all Fs < 2.65).

The last area of interest was the gap-site, i.e. the preposition in the subordinate clause. There were no main effects (all Fs < 1), but the interaction was significant by subjects ($F_1(1,31) = 7.41$, $p < 0.05$; $F_2 (1,15) = 2.52$, $p = 0.13$).

To sum up, the subordinator was read more slowly in the adverbial extraction condition compared to the complement extraction condition. Furthermore, since there was no difference in the non-extraction condition, this cannot be explained by the fact that there were systematic differences between the subordinators, e.g. the fact that the adverbial subordinator (*hvis*) is graphemically twice as long as the complementizer (*at*). It turned out that there were suggestions of an interaction at the gap-site, but this is very hard to interpret because the subjects had been experiencing difficulties with the adverbial extractions for a while.

2.3 Discussion of Experiments 1a and 1b

Experiment 1a confirmed that native speakers who are unaware of the issue at hand consider extraction out of adverbial clauses unacceptable, at least when the sentences have the specifications used in the experiments. Experiment 1b extended the conclusion by showing that native speakers react to the unacceptable sentences even when they are not asked to. Furthermore, the reading-time experiment replicated McKinnon & Osterhout's (1996) finding that difficulties in unacceptable extraction show up already at the transition from the main to the subordinate clause, rather than being delayed to the gap-site as suggested by Kluender & Kutas (1993).

Experiment 1a also suggested that not all complement clause extractions are equally acceptable. It has been suggested since Ross (1967)[5] that complement-taking verbs differ in the degree to which they allow extraction. Such differences in German have been experimentally investigated and found by Featherston (2004). The present study was not specifically designed to investigate this, so the present finding should be considered with some caution. But it seems that to the extent that matrix verbs influence the degree of acceptability of extraction, the effect is relatively small compared to the negative effect of extracting out of adverbial clauses, at least in the case of Danish.

3. EXPERIMENTS 2A AND 2B

With the results from the first set of experiments, we are now in a position to investigate non-syntactic factors that might influence extraction. Jensen (2001a), for

example, suggested that a high degree of semantic cohesion between the main and adverbial clauses would make extraction from the adverbial clause acceptable (cf. section 1.2). Semantic cohesion was defined in terms of whether the two clauses described different aspects of the same event or two separate events. Jensen gave as acceptable examples sentences where the main predicate encodes a psychological or physical state while the adverbial subordinate clause encodes a reason for the state. The idea was that cause and effect are aspects of the same complex event, and thus semantically cohesive.

The following two experiments test this idea by employing sentence pairs where the matrix predicate encoded a physical or psychological state, and where the subordinate clause could be construed as providing a reason for the state. As such, the sentences were set up to be maximally acceptable given Jensen's generalization over her authentic examples and intuitive judgments.

The sentences differed, however, with regard to how straightforwardly the subordinate clause could be interpreted as a reason for the state:

(13) De   saltstænger$_i$ bliver   jeg tørstig hvis jeg spiser __$_i$ inden aftensmaden.
       those pretzels$_i$ become I   thirsty if   I   eat __$_i$ before dinner
(14) De   julekort$_i$ bliver   jeg tørstig hvis jeg skriver __$_i$ inden
       those Christmas-cards$_i$ become I   thirsty if   I   write __$_i$ before
       aftensmaden.
       dinner

It is easier to construe the hypothetical eating event in (13), rather than the writing event in (14) as a reason for the physical state of being thirsty. When it is difficult to construe the subordinate clause as a reason for the state encoded in the matrix clause, I assume that there is a lower degree of semantic cohesion between them since it makes it more difficult to immediately construe the two clauses as aspects of the same complex cause–effect event. As a consequence, (13) should be rated more acceptable and processed with less difficulty than (14).

We can think about this theoretical prediction in a different way. In the cohesive condition, our world knowledge of being thirsty and what may elicit this state makes it possible to predict where the sentence is heading on the information available in the main clause. A number of researchers have in various ways suggested that a condition for extraction is that the entire sentence can be interpreted as being about the fronted element (Erteschik-Shir 1973, Allwood 1982:28, Kuno 1987, Van Valin 1995). In (13), it is consistently easier continually throughout the sentence to semantically connect each new piece of information to a proposition about the fronted element. In (14), on the other hand, it may be difficult to see what relevance becoming thirsty has in relation to Christmas cards.

3.1 Experiment 2a: acceptability

3.1.1 Procedure and participants

A total of 32 high school students from Stenløse Gymnasium and Ingrid Jespersens Gymnasium in Copenhagen participated on a voluntary basis. The mean age of this subset was 17;11, ranging from 16 years to 19 years.

The instructions and procedure were generally the same as in Experiment 1a with one exception. To increase the sensitivity of the measure, participants were asked to rate sentences for naturalness on a scale from 1 to 7 instead of 1 to 5 as in the previous experiment.

3.1.2 Design

The experiment used a 2 (cohesive vs. incohesive) by 2 (extraction vs. non-extraction) factorial design. The following sentences exemplify the different conditions:

Cohesive, extraction

(15) Den whisky bliver   hun fuld hvis hun bæller $\_i$ inden fodboldkampen.
      that whisky$_i$ becomes she drunk if   she quaffs $\_i$ before soccer.game-DEF

Incohesive, extraction

(16) Den kam$_i$ bliver   hun fuld hvis hun taber $\_i$ inden fodboldkampen.
      that comb$_i$ becomes she drunk if   she drops $\_i$ before soccer.game.DEF

Cohesive, non-extraction

(17) Hun bliver   fuld hvis hun bæller den whisky inden fodboldkampen.
      she becomes drunk if   she quaffs that whisky before soccer.game.DEF

Incohesive, non-extraction

(18) Hun bliver   fuld hvis hun taber den kam   inden fodboldkampen.
      she becomes drunk if   she drops that comb before soccer.game.DEF

As in the previous acceptability experiment, the non-extraction conditions were introduced because it is possible that semantic differences between the critical conditions are responsible for differences in acceptability ratings. In fact, since the incohesive sentences encode unlikely causal relationships, we would expect there to be substantial plausibility differences of a sort that could have an effect on acceptability judgments. This is obviously not optimal, but the idea with the non-extraction condition was that any effect due to semantic oddness could be determined and factored out on the basis of the non-extraction conditions.

3.1.3 Materials

The stimuli material contained 20 items, each consisting of four sentences representing each of the four experimental conditions. The entire set of stimuli sentences can be found in Appendix B. All sentences followed the same formula in the extraction conditions:

> extracted element + *bliver* 'become' + first person pronoun in nominative case + adjective + adverbial subordinator *hvis* 'if' + first person pronoun agreeing with the one in the matrix clause + transitive verb + temporal or locative PP

The extracted element was to be interpreted as the direct object of the subordinate transitive verb. It consisted of a demonstrative pronoun and an inanimate noun. The non-extraction sentences had the same specifications except that the subordinate direct object was in its canonical position after the subordinate verb and the matrix subject was positioned sentence-initially (due to V2).

Four counterbalanced lists were created, each containing exactly one sentence from each item and an equal number of sentences from the four different experimental conditions. Each list additionally contained 60 structurally unrelated filler sentences that represented an even spread of acceptability levels. Thus, each of the four lists contained 80 sentences. Each list was presented on a separate questionnaire in one of four different pseudo-random orders; there were always three fillers between each critical item, but the order within fillers and critical items respectively was random.

3.1.4 Predictions

If the semantic cohesion between clauses makes adverbial clause extraction acceptable, the cohesive extraction sentences should be rated more natural than the incohesive extraction sentences. Furthermore, we predict an interaction such that the difference between the cohesive and incohesive sentences should be greater between the extraction conditions than between the non-extraction conditions. The interaction is important because it is possible that participants will let the odd semantics of the incohesive conditions influence their acceptability ratings negatively. The interaction is thus necessary to show that cohesion has a special effect on extraction that goes beyond its general effect.

Jensen rated extractions from cohesive adverbial clauses acceptable. According to a literal interpretation of that, we should see no effect of extraction in the cohesive condition, only in the incohesive condition. But since Experiment 1a showed an effect of extraction even in extraction out of complement clauses, a more realistic prediction in Jensen's spirit would be that cohesion should interact with construction type such that cohesion should greatly ameliorate the negative effect of extraction.
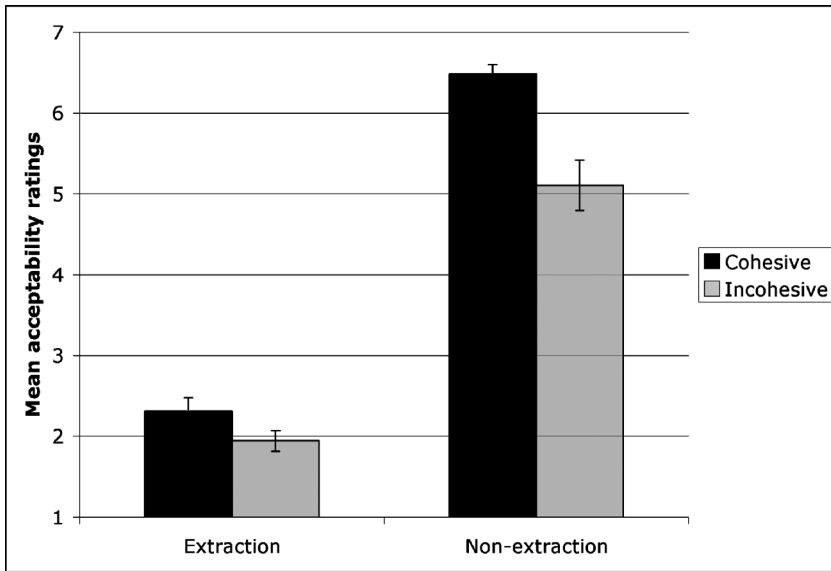
Figure 7. Mean acceptability ratings in Experiment 2a.

3.1.5 Results and discussion

The mean acceptability ratings are presented in Figure 7.

An analysis of variance revealed main effects of construction ($F_1(1,31) = 227.97$, $p < 0.001$; $F_2(1,19) = 1457.76$, $p < 0.001$) and cohesion ($F_1(1,31) = 22.82$, $p < 0.001$; $F_2(1,19) = 52.75$, $p < 0.001$), and an interaction ($F_1(1,31) = 19.86$, $p < 0.001$; $F_2(1,19) = 32.99$, $p < 0.001$). But as can be seen on the graph, the interaction went the wrong way; the cohesion factor mattered LESS in the extraction conditions than in the non-extraction conditions. This is the opposite of what was predicted by Jensen's semantic cohesion proposal.

The results can be interpreted in two ways, however. The most obvious conclusion is that both cohesion and extraction play a role for acceptability, but that extraction is a more decisive factor and that cohesion plays no special role for extraction. The reason why we see the interaction, then, is that if the sentence is nonsensical because extraction out of adverbial clauses is bad in the first place, then lack of cohesion is not going to make much of a difference.

The other possibility is that the difference between the extraction conditions was limited by a floor effect; maybe the incohesive extraction condition would have been rated much lower if there had been no bottom of the scale.

The second, and theoretically less interesting, interpretation is difficult to rule out, making it impossible to dismiss the possibility of a special effect of cohesion on

extraction. But we still find a rather dramatic effect of extraction even in the cohesive conditions: the mean rating shifts from 6.5 in the cohesive non-extraction condition to 2.3 in the cohesive extraction condition. This makes it difficult to maintain that semantic cohesion makes extraction out of adverbial clauses acceptable.

3.2 Experiment 2b: self-paced reading

The conscious task of judging the naturalness of a sentence is still an odd task that has little to do with the ordinary use of language. It is possible that the oddity of the task or misguided normative considerations influenced the results in experiment 2a. A follow-up reading-time experiment as in the first set of experiments was conducted to see if semantic cohesion would alleviate the difficulties of adverbial clause extraction in a task that did not require conscious reflection in the same way.

3.2.1 Design and materials

The same 2 (cohesion: cohesive vs. incohesive) by 2 (construction: extraction vs. non-extraction) factorial design as in Experiment 2a was used. The experiment also featured the same materials organized in the same counterbalanced and randomized stimuli lists.

3.2.2 Procedure and participants

Thirty-two students from various first-year university classes in the Faculty of Humanities, University of Copenhagen, were recruited to participate on a voluntary basis. The mean age was 23;8 years with a range from 19 years to 42 years.

The procedure was the same self-paced moving-window procedure as that in Experiment 1b.

3.2.3 Predictions

Given that the processing difficulties with extraction out of adverbial clauses were first observed at the subordinator in Experiment 1b, we would expect that the subordinator of the incohesive extraction condition would be read more slowly than the subordinators in the non-extraction conditions. If semantic cohesion makes extraction out of adverbial clauses possible, we should expect an alleviation of difficulties manifested as an interaction between construction and cohesion. The exact spot for the alleviation is not clear though. If extractions with semantic cohesion are to be processed completely without difficulties, we should be able to observe the interaction at the subordinator. This is somewhat counter-intuitive since the full nature of the cohesion has not been revealed. But it is possible that subjects would pick up on a greater likelihood for the fronted elements in the cohesive conditions to be causally linked to the matrix predicate as early as in the matrix clause. An alternative
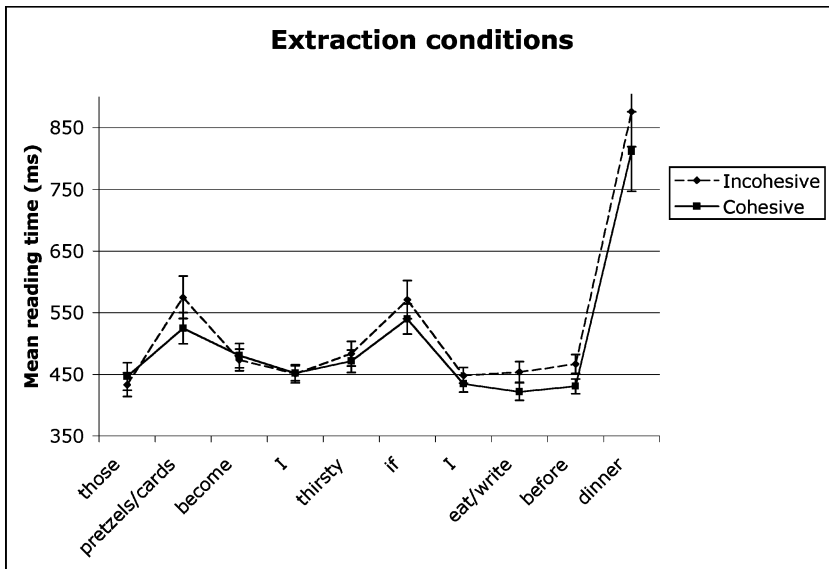
Figure 8. Experiment 2b mean reading times in extraction conditions.

possibility is that the effect does not show up until when the fronted element has to be integrated into the subordinate verb's argument structure. There has been and is some discussion in the psycholinguistic literature as to whether this happens at the verb or at the gap-site (e.g. Traxler & Pickering 1996, Hestvik et al. 2007, Phillips & Wagers 2007). Either way, we should see the interaction after the gap-site (at the preposition), assuming that the alleviating effect lasts more than one word. This is a reasonable assumption since the claim to be tested is that cohesion helps extraction to the point of acceptability. A last possibility is that the effect does not show up until the last word.

3.2.4 Results

As in Experiment 1b, data points above or below three standard deviations from the condition mean of a particular word position were replaced by the cut-off value. This affected 1.86 percent of the data. Figure 8 summarizes the data from the extraction conditions and Figure 9, the data from the non-extraction conditions.

The first point of theoretical interest is the subordinator. Figures 10 and 11 illustrates the reading times at the subordinator. Reading times were submitted to a 2 by 2 analysis of variance. The analysis revealed a strong main effect of construction ($F_1(1,31) = 17.80, p < 0.001$; $F_2(1,19) = 91.31, p < 0.001$), but only a marginal effect of cohesion in the by-subjects analysis ($F_1(1,31) = 2.95, p = 0.096$;
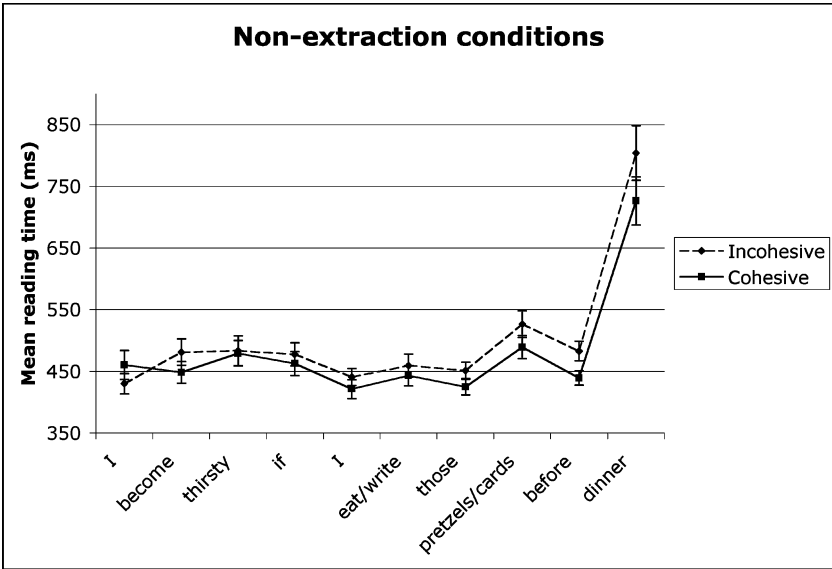
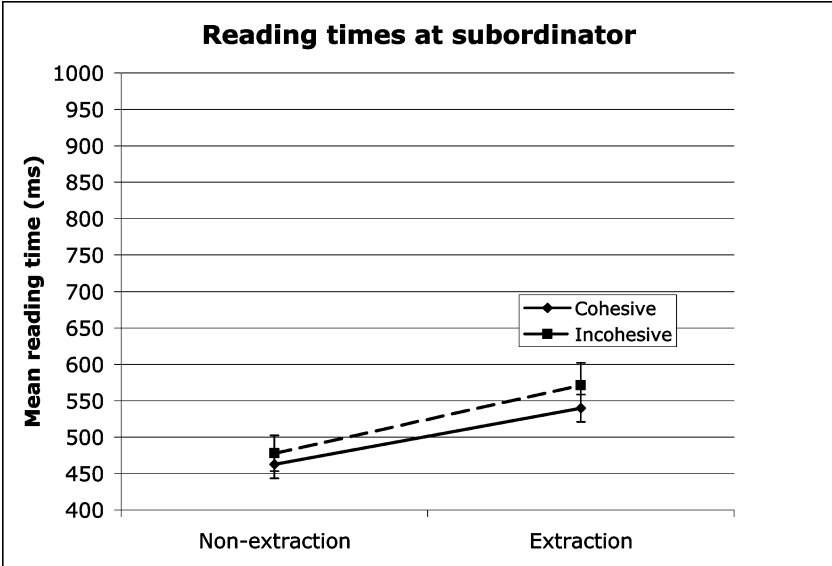**Figure 9.** Experiment 2b mean reading times in non-extraction conditions.

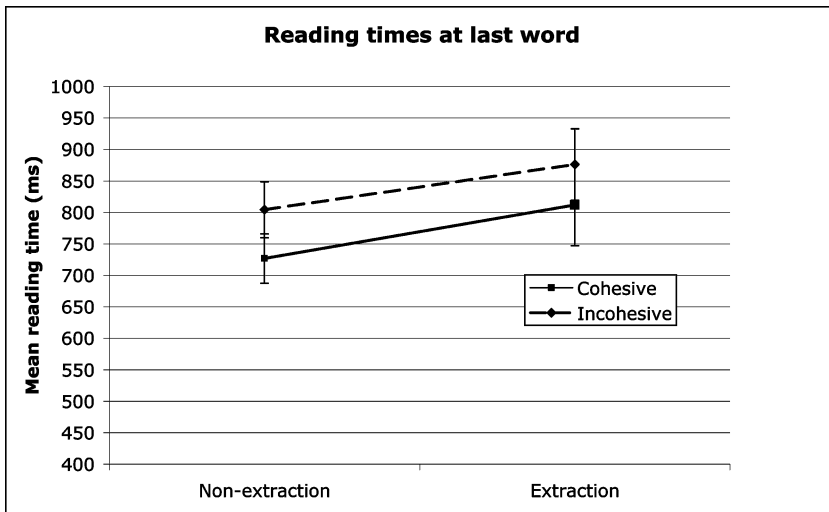**Figure 10.** Experiment 2b mean reading times at the subordinator.

Figure 11. Experiment 2b mean reading times at last word.

$F_2(1,19) = 1.25, p < 0.28$), and no interaction (both $Fs < 1$). At the next word, the subordinate subject, there were no significant effects, although the effect of cohesion approached significance by subjects ($F_1(1, 31) = 3.14, p = 0.09; F_2(1, 19) = 1.70, p = 0.20$). At the subordinate verb the effect of cohesion was significant by subjects ($F_1(1,31) = 11.99, p < 0.01; F_2(1,19) = 4.03, p = 0.06$). All other effects were insignificant ($Fs < 2.8$) at this position.

The next area of theoretical interest is the first word of the sentence-final PP, since this was the word that followed the gap in the extraction sentences. There was a main effect of cohesion ($F_1(1,31) = 27.02, p < 0.001; F_2(1,19) = 9.56, p < 0.01$). The effect of construction was marginally significant by items ($F_1(1, 31) = 2.56, p = 0.12; F_2(1,19) = 4.33, p = 0.051$). But there was no interaction between construction and cohesion (both $Fs < 1$).

Figure 11 shows the reading times at the last word. Analyses revealed a main effect of construction ($F_1(1,31) = 7.98, p < 0.01; F_2(1,19) = 5.11, p < 0.05$) and a main effect of cohesion in the by-subjects analysis ($F_1(1,31) = 8.35, p < 0.01$), but not by-items ($F_2(1,19) = 2.19, p = 0.16$). There was no interaction (both $Fs < 1$).

The general results of the experiment can be summarized as follows. Extraction caused early difficulties (at the subordinator), replicating the finding from Experiment 1b. Although cohesion did make processing easier at later points in the sentences, the effect applied equally to the extraction and non-extraction conditions. We can thus conclude that cohesion, or at least semantic plausibility, does influence processing, but that there is no evidence for the hypothesis that cohesion plays a special role in extraction.

3.3 General discussion of Experiments 2a and 2b

Seen together, the results from the two experiments point in the same direction. If the relationship between the two clauses in the stimuli sentences is incohesive, the sentences are rated very low on naturalness and are processed more slowly. This is not very surprising given the fact that the incohesive conditions were semantically odd. The point of interest of the experiments was whether cohesion would interact with extraction such that a high degree of cohesion would ease some of the difficulties associated with extraction out of adverbial clauses that were documented in Experiments 1a and 1b. Cohesion did to some degree influence the negative reaction to extraction out of adverbial clauses in the sense that participants reacted less negatively to the cohesive extraction condition than they did to the incohesive extraction condition in both experiments. But crucially this effect did not appear to be much different from the effect of cohesion seen on the non-extraction control conditions. The primary exception to this was the interaction effect found in Experiment 2a, but the interaction went against the prediction such that cohesion actually had less of an effect in the extraction conditions compared to the non-extraction controls. Two interpretations of that result were considered. One interpretation was that the general negative effect of lack of cohesion was made irrelevant by an unacceptable extraction. If the sentence is difficult to make heads or tails of because of structural or processing factors, the implausibility of the truth-conditional semantics should not make too much of a difference. The second interpretation was that the interaction was a consequence of the scale used being bounded and the crucial ratings lying at the bottom of the scale (a floor effect). It speaks against this interpretation that the original proposal was that a high degree of semantic cohesion should actually make extraction out of adverbial clauses acceptable. This was not the case. Although it is difficult, if not impossible, to impose an acceptable/unacceptable distinction on a seven-point scale in a motivated way, we still observe that in the cohesive conditions the mean naturalness rating shifted from the top of the scale in the non-extraction condition to near the bottom in the extraction condition. Furthermore, the results from Experiment 2b, which used a measure that did not suffer from the same limitations, provided no evidence for the right sort of interaction between construction type and cohesion.

One possible objection is that the cohesion manipulation used in the experiment was not optimal because it introduced an element of plausibility that, as we have observed, in general influenced the results significantly in a way that was irrelevant to the theoretical claim of Jensen. As stated before, this makes the non-extraction controls very important. But it is possible that it introduced unknown effects into the data that cannot simply be controlled for with the non-extraction sentences. The lack of interaction effect is not necessarily too damaging to the semantic cohesion account in this light because it is not a theory of plausibility as such. This is admittedly a

weakness of the design of the two experiments; it subtracts from the experiments' ability to falsify the semantic cohesion constraint.

Overall, however, the conclusion that a high degree of semantic cohesion does not make extraction out of adverbial clauses acceptable still appears valid. The dramatic effect of extraction on acceptability ratings even in the cohesive conditions is completely independent of whatever faults can be attributed to cohesive/incohesive manipulation. Most importantly the cohesive extraction sentences followed the semantic specifications of Jensen and were almost structurally identical to Jensen's own examples such as:

(19) Det$_i$ bliver   jeg glad hvis du vil   gøre_$_i$.
     that$_i$ become I   glad if   you will do _$_i$ (Jensen 2001a:25, ex. (11))

So even though such sentences may be produced occasionally, they do seem to cause difficulties for the comprehender.

The general conclusion from the two experiments thus seems to be that extraction out of adverbial clauses is unacceptable and causes difficulties even when there is a high degree of semantic cohesion between the two clauses, at least under the conditions in the present experiments. Although there was plenty of evidence that cohesion or semantic plausibility influences both acceptability and reading time, no evidence was found that this effect had any special ties to extraction.

4. GENERAL DISCUSSION

Through the years much has been said in the linguistics literature about the constraints on extraction, but with a few exceptions, most of the theorizing has been done on uncertain empirical grounds. This is not to say that we should throw away all we know about extraction. The possibility of extraction out of relative clauses in Danish and Swedish seems fairly well-documented with authentic examples (Engdahl 1997; Jensen 1998, 2001a, b), but authentic examples are difficult to use in teasing out what IS NOT possible and why.

Experimental methodologies allow linguists to quantify the degree to which sentences are acceptable or unacceptable (under a certain set of conditions) and to intersubjectively test hypotheses about factors that are thought to influence acceptability. The main purpose of the first set of experiments in this paper was to establish some basic facts about the distinction between complement and adverbial clause extraction in Danish with regard to how they are rated and processed. Experiment 1a confirmed the general notion that complement clause extraction is acceptable and that the investigated adverbial clause extractions are unacceptable. Experiments 1b and 2b showed that extraction from adverbial clauses is associated with a processing cost from the subordinator on, mirroring McKinnon &

Osterhout's (1996) previous ERP findings in English. This suggests that a similar constraint is at work across the two languages. It should be noted, though, that none of the studies seriously addresses what the nature of the constraint is, i.e. whether it is syntactic, semantic or pragmatic in nature (cf. Poulsen 2005 for speculations on this subject).

Experiment 1a furthermore suggested that even in generally acceptable complement clause extraction, verbs differ in the degree to which they allow extraction. This is by no means a new idea (Ross 1967), but by using quantitative measures it is possible to say how much choice of matrix verb affects acceptability relative to constraint violations such as extraction from adverbial clauses. Considering the verbs used in the present experiment the answer is: not a lot. But a fairly limited range of verbs was used. It has been suggested that the semantic heaviness of a verb is the controlling factor (e.g. Erteschik-Shir & Lappin 1979), but it is a matter of future research to substantiate this claim more thoroughly.

The second set of experiments investigated the proposal that extraction is possible out of certain adverbial clause in Danish. Specifically, Jensen (2001a) has suggested that semantic cohesion between the main and adverbial clauses makes extraction out of the adverbial clause acceptable. Experiments 2a and 2b confirmed the findings from the first set of experiments that extraction out of an adverbial clause is associated with a steep acceptability and processing cost compared to similar sentences with no extraction, but there was no indication that semantic cohesion plays a special role in extraction. The semantically cohesive sentences that were rated near the top of the scale without extraction dropped to the bottom of the scale with extraction, and there was no indication in the reaction-time data that semantic cohesion influenced the processing of extraction sentences over and above the general effect that reading something cohesive is easier than reading something that is not.

The linguistic conclusion is that at present there is limited empirical reason to include the notion of semantic cohesion in the grammatical description of extraction in Danish. Whether extraction out of adverbial clauses in general should be considered outside Danish grammar is still somewhat open, but at the moment we do not have consistent evidence to suggest that extraction from adverbial clauses is possible, while we do have evidence that at least some extractions – INCLUDING ONES WITH THE STRUCTURAL AND SEMANTIC SPECIFICATION OF THE RARE AUTHENTIC EXAMPLES – are rated unacceptable and cause processing difficulty. The rare examples encountered should probably be cautiously considered speech errors, anomalies or the result of individual idiosyncrasies. If considered speech errors, they could still be motivated by the tendency to front in Danish. The reason for considering them errors nevertheless would then be that they are problematic for the listeners.

The use of both acceptability judgment and reading-time measures were motivated by a concern for the validity of any one experimental measure. The task of acceptability judgment is somewhat suspect, and there is good reason to doubt that

it is possible to get informants to do exactly what they are supposed to do (Schütze 1996, Poulsen 2005). But by demonstrating effects using different methodologies, we reduce the risk of accepting effects that are artifacts of the methodology used. The present results using two different methodologies supplement each other, and both of the reading-time experiments found prolonged processing at the subordinator of sentences with extraction out of adverbial clauses, which as mentioned replicates previous findings (McKinnon & Osterhout 1996).

One valid objection that can be raised is that the experimental methodologies used and cited here did not investigate extraction in context. Context and pragmatic factors have long been claimed to be the primary determinant of the acceptability of extraction in Scandinavian linguistics (e.g. Erteschik-Shir 1973; 1982; Erteschik-Shir & Lappin 1979; Allwood 1982; Engdahl 1997). Investigating contextual effects is, however, difficult because a simple demonstration that sentences are rated higher when context is provided could be interpreted in a number of ways unless the context completely removes any negative effect of extraction. Allwood (1982:26) suggested that the Swedish relative clause extraction in (20) is easier to accept if read in the context of (21):

(20) En sköldpadda$_i$ känner jag en flicka som kyssta _$_i$.
       a tortoise$_i$ know I a girl who kissed _$_i$
(21) Känner du någon   som kysst et djur?
       know   you anybody who kissed an animal

The suggested reason was that the question makes it easier to interpret the extraction sentence as being about the extracted element. An alternative explanation is possible, however, because the context question provides information about who we should be expect to be doing what whom, i.e. that the speaker is the first argument of the knowing, and that a person is the first argument and an animal the second argument of the kissing. In other words, the argument relations of the extraction sentence can be guessed from the context and lexical semantics WITHOUT KNOWING ANYTHING ABOUT SWEDISH GRAMMAR. As we saw in Experiments 2a and 2b, sentences with easy-to-understand semantics are rated higher and processed faster than sentences that encode difficult-to-understand semantics, so differences in responses to (20) depending on presence or absence of (21) could be interpreted as reflecting semantic redundancy rather than facts about the grammar of extraction.

This is not to say that context does not play a role in extraction, just that demonstrating it empirically is somewhat more complex than simply registering a positive effect of context. One possible avenue to explore would be to investigate whether contexts that specifically fit the discourse function of extraction show alleviating effects compared to contexts that do not.

The present paper has only dealt with extraction out of adverbial clauses. The Scandinavian languages are probably more famed for the possibility of extraction out

of relative clauses as mentioned in the introduction. The evidence for the possibility of this type of extraction seems far less uncertain than extraction out of adverbial clauses, so here the task ahead instead is to test theories of WHEN extraction out of relative clauses is possible (e.g. Erteschik-Shir 1973, 1982).

For the sake of simplicity, context and relative clauses were left out of the present experiments, but systematic investigations of these issues are obvious subjects of future studies. Such studies are possible now that we have baseline ideas about how people react to extraction constructions.

APPENDIX

A. Materials for Experiments 1a and 1b

The 16 items of Experiments 1a and 1b are presented below. The words before the slash were used in the complement clause conditions, while the words after the slash were used in the adverbial conditions. The numbers signify the item number. Extraction conditions are marked with *e*, non-extraction conditions are marked with *ne*.

1e Den mappe tror/besvimer jeg at/hvis de glemmer på kontoret.
that briefcase believe/faint I that/if they forget at office.DEF

1ne Jeg tror/besvimer at/hvis de glemmer den mappe på kontoret.
I believe/faint that/if they forget that briefcase at office.DEF

2e Den øl tror/besvimer jeg at/hvis jeg drikker på ti minutter.
that beer believe/faint I that/if I drink in ten minutes

2ne Jeg tror/besvimer at/hvis jeg drikker den øl på ti minutter.
I think/faint that/if I drink that beer in ten minutes

3e Det stof mener/griner jeg at/hvis hun vælger til gardinerne.
that fabric believe/laugh I that/if she pick for drapes.DEF

3ne Jeg mener/griner at/hvis hun vælger det stof til gardinerne.
I believe/laugh that/if she pick that fabric for drapes.DEF

4e Den cykel mener/griner jeg at/hvis han køber på auktionen hos politiet.
that bike believe/laugh I that/if he buy at auction.DEF at police.DEF

4ne Jeg mener/griner at/hvis han køber den cykel på auktionen hos politiet.
I believe/laugh that/if he buy that bike at auction.DEF at police.DEF

5e Den kagedåse ved/græder jeg at/hvis du gemmer på den øverste hylde.
that cookie.jar know/cry I that/if you hide on the top shelf

5en Jeg ved/græder at/hvis du gemmer den kagedåse på den øverste hylde.
I know/cry that/if you hide that cookie.jar on the top shelf

6e De penge ved/græder jeg at/hvis han stjæler fra min frakkelomme.
that money know/cry I that/if he steal from my coat.pocket

6ne Jeg ved/græder at/hvis han stjæler de penge fra min frakkelomme.
I know/cry that/if he steal that money from my coat.pocket

7e Det referat erkender/hyler jeg at/hvis de fremlægger på onsdag.
that report admit/cry.out I that/if they present on Wednesday

7ne Jeg erkender/hyler at/hvis de fremlægger det referat på onsdag.
I admit/cry.out that/if they present that report on Wednesday

8e Den lov erkender/hyler jeg at/hvis de vedtager i Folketinget.
that law admit/cry.out I that/if they pass in parliament.DEF

8ne Jeg erkender/hyler at/hvis de vedtager den lov i Folketinget.
I admit/cry.out that/if they pass that law in parliament.DEF

9e Det billede påstår/rødmer jeg at/hvis hun udstiller på museet i Århus.
that painting claim/blush I that/if she exhibit at museum.DEF in Århus

9ne Jeg påstår/rødmer at/hvis hun udstiller det billede på museet i Århus.
I claim/blush that/if she exhibit that painting at museum.DEF in Århus

10e Den fejl påstår/rødmer jeg at/hvis de opdager i regnskabet.
that error claim/blush I that/if they detect in account.DEF

10ne Jeg påstår/rødmer at/hvis de opdager den fejl i regnskabet.
I claim/blush that/if they detect that error in account.DEF

11e Den kylling formoder/dør jeg at/hvis jeg spiser til aftensmad.
that chicken presume/die I that/if I eat for dinner

11ne Jeg formoder/dør at/hvis jeg spiser den kylling til aftensmad.
I presume/die that/if I eat that chicken for dinner

12e Den dessert formoder/dør jeg at/hvis du serverer efter osten.
that dessert presume/die I that/if you serve after cheese.DEF

12ne Jeg formoder/dør at/hvis du serverer den dessert efter osten.
I presume/die that/if you serve that dessert after cheese.DEF

13e Det tæppe konkluderer/nyser jeg at/hvis han ryster inde i stuen.
that blanket conclude/sneeze I that/if he shakes inside in living.room.DEF

13ne Jeg konkluderer/nyser at/hvis han ryster det tæppe inde i stuen.
I conclude/sneeze that/if he shakes that blanket inside in living.room.DEF

14e Det krydderi konkluderer/nyser jeg at/hvis de bruger i maden.
that spice conclude/sneeze I that/if they use in food.DEF

14ne Jeg konkluderer/nyser at/hvis de bruger det krydderi i maden.
I conclude/sneeze that/if they use that spice in food.DEF

15e Den pude håber/snorker jeg at/hvis du fjerner fra min seng.
that pillow hope/snore I that/if you remove from my bed

15ne Jeg håber/snorker at/hvis du fjerner den pude fra min seng.
I hope/snore that/if you remove that pillow from my bed

16e Det vindue håber/snorker jeg at/hvis du lukker i nat.
that window hope/snore I that/if you close in tonight

16ne Jeg håber/snorker at/hvis du lukker det vindue i nat.
I hope/snore that/if you close that window in tonight

B. Materials for Experiments 2a and 2b

The 20 items of Experiments 2a and 2b are presented below. The words before the slash were used in the cohesive conditions, while the words after the slash were used in the incohesive conditions. The numbers signify the item numbers. Extraction conditions are marked with *e*, non-extraction conditions are marked with *ne*.

1e De saltstænger/julekort bliver jeg tørstig hvis jeg spiser/skriver inden aftens-maden.
those pretzels/Christmas cards become I thirsty if I eat/write before dinner

1ne Jeg bliver tørstig hvis jeg spiser/skriver de saltstænger/julekort inden aftens-maden.
I become thirsty if I eat/write those pretzels/Christmas cards before dinner

2e Den lasagne/tavlesvamp bliver hun mæt hvis hun æder/gemmer efter under-visningen.
that lasagna/eraser become she full if she eat/hide after class.DEF

2ne Hun bliver mæt hvis hun æder/gemmer den lasagne/tavlesvamp efter under-visningen.
she become full if she eat/hide that lasagna/eraser after class.DEF

3e Den whisky/kam bliver hun fuld hvis hun bæller/taber inden fodboldkampen.
that whiskey/comb become she drunk if she quaff/drop be-fore the soccer.match.DEF

3ne Hun bliver fuld hvis hun bæller/taber den whisky/kam inden fodboldkampen.
she become drunk if she quaff/drop that whiskey/comb before soccer.match.DEF

4e Den tur/væg bliver jeg udmattet hvis jeg løber/betragter før mødet.
that lap/wall become I exhausted if I run/look.at before meeting.DEF

4ne Jeg bliver udmattet hvis jeg løber/betragter den tur/væg før mødet.
I become exhausted if I run/look.at that lap/wall before meeting.DEF

5e Den jakke/kande bliver du kold hvis du åbner/vasker på udflugten.
that coat/jug become you cold if you open/wash on excursion.DEF

5ne Du bliver kold hvis du åbner/vasker den jakke/kande på udflugten.
you become cold if you open/wash that coat/jug on excursion.DEF

6e De vitaminer/bogreoler bliver du sund hvis du tager/tømmer hver morgen.
those vitamins/book-cases become you healthy if you take/empty every morning

6ne Du bliver sund hvis du tager/tømmer de vitaminer/bogreoler hver morgen.
you become healthy if you take/empty those vitamins/book.cases every morning

7e Den læbestift/hoveddør bliver hun lækker hvis hun bruger/lukker inden festen.
that lipstick/front.door become she lovely if she use/close before party.DEF

7ne Hun bliver lækker hvis hun bruger/lukker den læbestift/hoveddør inden festen.
she become lovely if she use/close that lipstick/front.door before party.DEF

8e Den øl/sten bliver han søvnig hvis han drikker/kaster under receptionen.
that beer/stone become he sleepy if he drink/throw during reception.DEF

8ne Han bliver søvnig hvis han drikker/kaster den øl/sten under receptionen.
he become sleepy if he drink/throw that beer/stone during reception.DEF

9e Den frituregryde/det nyhedsmagasin bliver han forbrændt hvis han vælter/køber på torsdag.
that.COM deep.fryer/that.NEUT news.magazine become he burnt if he overturn/buy on Thursday

9ne Han bliver forbrændt hvis han vælter/køber den frituregryde/det nyhedsmagasin på torsdag.
he become burnt if he overturn/buy that.COM deep.fryer/that.NEUT news. magazine on Thursday

10e Den kur/båd bliver du slank hvis du følger/døber efter nytår.
that diet/boat become you slim if you follow/baptize after new.year

10ne Du bliver slank hvis du følger/døber den kur/båd efter nytår.
you become slim if you follow/baptize that diet/boat after new.year

11e Det bjerg/skrin bliver han svimmel hvis han bestiger/begraver i februar.
that mountain/chest become he dizzy if he climb/burry in February

11ne Han bliver svimmel hvis han bestiger/begraver det bjerg/skrin i februar.
he become dizzy if he climb/burry that mountain/chest in February

12e Den parasol/madkasse bliver du solskoldet hvis du fjerner/husker i eftermiddag.
that parasol/lunchbox become you sun.scorched if you remove/remember this afternoon

12ne Du bliver solskoldet hvis du fjerner/husker den parasol/madkasse i eftermiddag.
you become sun.scorched if you remove/remember that parasol/lunchbox this afternoon

13e Den paraply/bordlampe bliver jeg gennemblødt hvis jeg glemmer/slukker hos Mormor.
that umbrella/desk.lamp become I soaked if I forget/turn.off at grandmother (grandmother's place)

13ne Jeg bliver gennemblødt hvis jeg glemmer/slukker den paraply/bordlampe hos Mormor.
I become soaked if I forget/turn.off that umbrella/desk.lamp at grandmother (grandmother's place)

14e Den bog/sok bliver du klog hvis du studerer/stopper på biblioteket.
that book/sock become you knowledgeable if you study/darn in library.DEF

14ne Du bliver klog hvis du studerer/stopper den bog/sok på biblioteket.
you become knowledgeable if you study/darn that book/sock in library.DEF

15e Det bål/skab bliver hun varm hvis hun tænder/henter i nat.
that bonfire/closet become she warm if light/fetch in tonight

15ne Hun bliver varm hvis hun tænder/henter det bål/skab i nat.
she become warm if she light/fetch that bonfire/closet in tonight

16e   Den skorsten/musik bliver du snavset hvis du renser/hører hos slagteren.
that chimney/music become you dirty if you clean/hear at butcher.DEF

16ne   Du bliver snavset hvis du renser/hører den skorsten/musik hos slagteren.
you become dirty if you clean/hear that chimney/music at butcher.DEF

17e   Den cigar/taske bliver hun dårlig hvis hun ryger/sælger efter fødselsdagen.
that cigar/bag become she queasy if she smoke/sell after birthday.DEF

17ne   Hun bliver dårlig hvis hun ryger/sælger den cigar/taske efter fødselsdagen.
she become queasy if she smoke/sell that cigar/bag after birthday.DEF

18e   Det halstørklæde/den glastallerken bliver jeg forkølet hvis jeg mister/smadrer inden skituren.
that.NEUT scarf/that.COM glass.plate become I cold (catch a cold) if I loose/break before ski.trip.DEF

18ne   Jeg bliver forkølet hvis jeg mister/smadrer det halstørklæde/den glastallerken inden skituren.
I become cold (catch a cold) if I loose/break that.NEUT scarf/that.COM glass.plate before ski.trip.DEF

19e   De diamanter/frikadeller bliver hun rig hvis hun stjæler/steger under udstillingen.
those diamonds/meat.balls become she rich if she steal/fry during exhibition.DEF

19ne   Hun bliver rig hvis hun stjæler/steger de diamanter/frikadeller under udstillingen.
she become rich if she steal/fry those diamonds/meat.balls during exhibition.DEF

20e   De strækøvelser/kaffekopper bliver du smidig hvis du laver/bytter efter arbejde.
those stretching.excercises/coffee.cups become you supple if you do/exchange after work

20ne   Du bliver smidig hvis du laver/bytter de strækøvelser/kaffekopper efter arbejde.
you become supple if you do/exchange those stretching.excercises/coffee.cups after work

ACKNOWLEDGEMENTS

NOTES

1. However, it should be noted that Taraldsen (1982) argued that Norwegian relative clause extraction is not a counterexample to the Complex NP constraint.

2. Example produced in conversation (from Engdahl 1997).
3. Danish is a V2 language and thus the verb is always in the second position. If a non-subject element is positioned preverbally, as in the extraction sentences, the subject is positioned after the verb.
4. This extractability score is similar in spirit to Featherston's (2004) 'bridgeness' score. The difference is that whereas Featherston's score is simply computed from how extraction sentences were rated, the present score takes into account that the combination of specific verbs and sentence contexts may influence acceptability regardless of extractability. In relation to the present experiment, this is clearly the appropriate measure because the sentences and the verbs were not always perfectly matched for plausibility due to the constraints on material construction. Whether similar factors are present in Featherston's materials, I cannot say, but the constraints in his materials seemed less likely to create the problems found in the present design.
5. Ross (1967:138f.) attributes the observation to Janet Dean Fodor.

REFERENCES

Allwood, Jens. 1982. The complex NP constraint in Swedish. In Engdahl & Ejerhed (eds.), 15–32.

Anderson, Lars-Gunnar. 1982. What is Swedish an exception to? Extractions and island constraints. In Engdahl & Ejerhed (eds.), 33–45.

Bard, Ellen G., Dan Robertson & Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72, 32–68.

Bock, Kathryn & Wilhelm Levelt. 1994. Language production. Grammatical encoding. In Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*, 945–984. San Diego, CA: Academic Press.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart and Winston.

Cohen, Jonathan, Brian MacWhinney, Matthew Flatt & Jefferson Provost. 1993. PsyScope: A new graphic interactive environment for designing psychological experiments. *Behavioral Research Methods, Instruments, and Computers* 25, 257–271.

Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.

Creider, Chet A. 1987. Structural and pragmatic factors influencing the acceptability of sentences with extended dependencies in Norwegian. *University of Trondheim Working Papers in Linguistics* 4, 2–141.

Dąbrowska, Ewa. 1997. The LAD goes to school. *Linguistics* 35, 735–766.

Deane, Paul. 1991. Limits to attention: A cognitive theory of island constraint phenomena. *Cognitive Linguistics* 2, 1–63.

Devriendt, Betty, Louis Goossens & Johan Van Der Auwera (eds.). 1995. *Complex structures: A functionalist perspective*. Berlin: Mouton de Gruyter.

Engdahl, Elisabet. 1982. Restrictions on unbounded dependencies in Swedish. In Engdahl & Ejerhed (eds.), 151–174.

Engdahl, Elisabet. 1997. Relative clause extraction in context. *Working Papers in Scandinavian Syntax* 60, 51–79.

Engdahl, Elisabet & Eva Ejerhed (eds.). 1982. *Readings on unbounded dependencies in Scandinavian languages*. Stockholm: Almqvist & Wiksell International.

Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Ph.D dissertation, MIT.

Erteschik-Shir, Nomi. 1982. Extractability in Danish and the pragmatic principle of dominance. In Engdahl & Ejerhed (eds.), 175–191.

Erteschik-Shir, Nomi & Shalom Lappin. 1979. Dominance and the functional explanation of island constraints. *Theoretical Linguistics* 6, 43–84.

Featherston, Sam. 2004. Bridge verbs and V2 verbs – the same thing in spades. *Zeitschrift für Sprachwissenschaft* 23, 181–209.

Hestvik, Arild, Nathan Maxfield, Richard G. Schwartz & Valerie Shafer. 2007. Brain responses to filled gaps. *Brain and Language* 100, 301–316.

Jakobsen, Lisbeth Falster. 1995. Sentence intertwining in Danish, seen from a Functional Grammar perspective. In Devriendt et al. (eds.), 61–92.

Jensen, Anne. 1998. *Knudekonstruktioner – en syntaktisk, semantisk og pragmatisk analyse af sætningsknuder i dansk*. MA thesis, University of Copenhagen.

Jensen, Anne. 2001a. Sentence intertwining in Danish. In Elisabeth Engberg-Pedersen & Peter Harder (eds.), *Ikonicitet og struktur*, 23–39. Preprint from Netværk for Funktionel Lingvistik, Department of English, University of Copenhagen.

Jensen, Anne. 2001b. Sætningsknuder – og Role and Reference Grammar. In Peter Widell & Mette Kunøe (eds.), *8. Møde i Udforskningen af Dansk Sprog*, 156–165. Aarhus Universitet, 12–13 October 2000.

Just, Marcel A., Patricia A. Carpenter & Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 11, 228–238.

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In Peter W. Culicover & Louise McNally (eds.), *The limits of syntax*, 241–279. San Diego, CA: Academic Press.

Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomena. *Language and Cognitive Processes* 8, 573–633.

Kuno, Susumu. 1987. *Functional Syntax*. Chicago, IL: University of Chicago Press.

Labov, William. 1975. Empirical foundations of linguistic theory. In Robert Austerlitz (ed.), *The scope of American linguistics*, 77–133. Lisse: Peter de Ridder Press.

McKinnon, Richard & Lee Osterhout. 1996. Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. *Language and Cognitive Processes* 11, 495–523.

Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82, 795–823.

Phillips, Colin & Matthew Wagers. 2007. Relating structure and time in linguistics and psycholinguistics. In Gareth Gaskell (ed.), *Oxford handbook of psyholinguistics*, 739–756. Oxford: Oxford University Press.

Poulsen, Mads. 2005. *Sentence processing and grammaticality in functional linguistics: Extraction in Danish as an example*. Ph.D. dissertation, University of Copenhagen.

Ross, John Robert. 1967. *Constraints on variables in syntax*. Ph.D dissertation, MIT.

Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago, IL: University of Chicago Press.

Snow, Catherine E. & Guus Meijer. 1977. On the secondary nature of syntactic intuitions. In Sidney Greenbaum (ed.), *Acceptability in language*, 163–177. The Hague: Mouton.

Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31, 575–582.

Taraldsen, Knut T. 1982. Extraction from relative clauses in Norwegian. In Engdahl & Ejerhed (eds.), 205–221.

Togeby, Ole. 2003. *Fungerer denne sætning: Funktionel dansk sproglære*. København: Gads Forlag.

Traxler, Matthew & Martin J. Pickering. 1996. Plausability and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35, 454–475.

Van Valin, Robert D. 1995. Towards a functionalist account of so-called extraction constraints. In Devriendt et al. (eds.), 29–60.

Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.