Not all processing difficulties are created equal

Johannes Kizach
Lionbridge

Abstract
A slowdown in reaction time in self-paced reading is typically interpreted as a sign of processing difficulty. Similarly, a low acceptability rating can, among other things, be caused by processing difficulty. The question examined in this article is whether a slowdown in reaction time always affects acceptability negatively. To investigate this, an acceptability study was performed, comparing sentences that only differ regarding the main verb (Mia noticed/presumed the pig in the pen needed water) and where the word-for-word reaction time data in reading is known from previous research. The reaction time data show that both types involve a slowdown, but at different locations in the sentence (at the embedded subject vs. at the embedded verb) and for different reasons (missing complementizer vs. reanalysis). The acceptability ratings show that the two types of slowdowns are not equally costly: The slowdown due to reanalysis causes a significantly lower rating than the slowdown due to a missing complementizer. The result illustrates that not all processing difficulties (measured as a slowdown in self-paced reading) have the same adverse impact on acceptability judgments.

1. Introduction
In self-paced reading experiments reaction time (RT) is measured and a slowdown is typically interpreted as a sign of increased processing difficulty (cf. Aaronson & Scarborough 1976; Baayen & Milin 2010; Cai, Sturt & Pickering 2012; Jegerski 2014). In acceptability judgment experiments, it is assumed that low acceptability can be caused by several factors, including ungrammaticality, low frequency of the lexical items, and processing difficulty (Schütze 2016: 160–162). The question is if all slowdowns lead
to reduced acceptability or if there are “cost free” processing difficulties: slowdowns in on-line processing that do not adversely affect acceptability.¹

Let us first look at a number of examples demonstrating the link between processing difficulty (slowdown) and reduced acceptability. Kizach & Balling (2013) conducted a speeded acceptability study where double object constructions with indefinite-definite and definite-indefinite order as in (1) and (2) were compared.

1

I call an RT-slowdown which does not result in reduced acceptability ‘cost free’, but clearly there is a cost since RT is slower. In this article I use cost free processing difficulty in this way instead of the more precise, but cumbersome term ‘processing difficulty which is not reflected in acceptability scores’.

(1) **Danish**

Direktør Clausen lovede [manden] [et arbejde]. [DEF-INDEF]

*President Clausen promised man-the a job*

‘President Clausen promised the man a job.’

(2) **Danish**

Direktør Clausen lovede [en mand] [arbejdet]. [INDEF-DEF]

*President Clausen promised a man job-the*

‘President Clausen promised a man the job.’

Based on similar studies of English (Brown, Savova & Gibson 2012; Clifton & Frazier 2004), the examples with indefinite-definite order (INDEF-DEF) were expected to be harder to process, and indeed, the RT was around 600 ms slower on average for the INDEF-DEF condition, illustrated in (2), compared to the DEF-INDEF condition, illustrated in (1) (Kizach & Balling 2013: 1162). Interestingly, the rejection rate (participants either accepted or rejected the sentences) was 16% for the INDEF-DEF condition, but only 4% for the DEF-INDEF condition (Kizach & Balling 2013: 1164). Despite the fact that both orders (definite-indefinite and indefinite-definite) are perfectly grammatical in Danish, participants decided to reject INDEF-DEF sentences in 16% of the cases – this appears to be an example of processing difficulty affecting acceptability.

Another example is from Christensen, Kizach & Nyvad (2013) where long and short *wh*-movement, as in (3) and (4) below, were compared in an acceptability judgment experiment.
(3) **Danish**

Hvad ved hun godt at man kan leje dér?

*What knows she well that one can rent there*

‘What does she know that one can rent there?’

(4) **Danish**

Ved hun godt hvad man kan leje dér?

*She knows well what one can rent there*

‘Does she know what one can rent there?’

Various processing theories predict that the longer the distance between a *wh*-element and a gap is, the harder processing should become (cf. Gibson 1998; Gibson 2000; Gibson 2003; Hawkins 1994; Hawkins 2004). In other words (3) should be harder to process than (4). Both are completely grammatical Danish sentences, but nevertheless (4) was judged to be significantly more acceptable than (3) – mean acceptability (on a 1 to 5 scale) was 4.76 and 3.66 respectively (Christensen, Kizach & Nyvad 2013: 57). Again, this is an example of processing difficulty having an adverse effect on acceptability.

In Fanselow & Frisch (2006), the contrast between (5) and (6) was investigated, and it turned out that (5) was judged to be better than (6) even though both are grammatical in German.

(5) **German** (Fanselow & Frisch 2006: 312 (15.21a))

Was denken Sie, dass die Entwicklung beeinflusst hat?

*What think you that the development influenced has*

‘What do you think influences the development?’

(6) **German** (Fanselow & Frisch 2006: 312 (15.21b))

Wer denken Sie, dass die Entwicklung beeinflusst hat?

*Who think you that the development influenced has*

‘Who do you think influences the development?’

Fanselow & Frisch (2006) interpreted this result as evidence for processing difficulties *increasing* acceptability. Their reasoning was that in (5) there is a temporary misanalysis because the reader initially takes *was* ‘what’ to be the object of *denken* ‘think’, but as soon as the embedded verb *beeinflusst* ‘influenced’ is parsed, the correct analysis is reached (where *was* is the object of *beeinflusst*, not *denken*). In (6) on the other hand *wer* ‘who’ is
nominative and cannot be analyzed as the object of denken ‘think’, which means that there is no temporary misanalysis of wer ‘who’. The higher acceptability of (5) could then suggest that the temporary misanalysis of was ‘what’ actually increases the acceptability. The reason for this facilitating effect of misanalysis is, according to Fanselow & Frisch (2006: 312–313), that the parser prefers as short a distance as possible between the wh-element and the gap – and during the temporary misanalysis, the wh-element was ‘what’ is closer to the gap than was ‘what’ and wer ‘who’ ultimately turn out to be.

I will note that there is another possible interpretation of this result, which is that the parser follows the Active Filler strategy (Frazier & Clifton Jr 1989; Frazier & d’Arcais 1989) or the Attach Anyway principle (Fodor & Inoue 1998). Both would lead to the result that the wh-element is attached as the object of denken ‘think’ – the difference would be that was ‘what’ is an acceptable object, whereas the attachment of wer ‘who’ leads to an anomaly. This temporary anomaly is then the cause of the reduced acceptability. Interpreted this way the results from Fanselow & Frisch (2006) are not evidence for a facilitating effect of processing difficulty, but another example of the adverse effect of processing difficulties.

In an attempt to determine precisely which of these two explanations is the right one, Kizach, Nyvad & Christensen (2013: 3) looked at the following contrast:

(7) Danish
Hvilket sprog har studenten lært at tale?
*Which language has student-the learnt to speak*
‘Which language has the student learnt to speak?’

(8) Danish
Hvilket afløb har studenten lært at rense?
*Which drain has student-the learnt to clean*
‘Which drain has the student learnt to clean?’

The wh-element in (7), hvilket sprog ‘which language’, is a plausible object for the matrix verb, lært ‘learnt’, but the wh-element in (8), hvilket afløb ‘which drain’, is not. If only the plausible wh-element is temporarily misanalysed as the object of the matrix verb, then we would expect reanalysis at the embedded verb only in the plausible condition. If on the
other hand both *wh*-elements are temporarily misanalysed, we would ex-
pect reanalysis in both conditions. Kizach, Nyvad & Christensen (2013:
3–4) conducted a self-paced reading experiment and found an RT-differ-
ence at the matrix verb, *lært* ‘learnt’, where (8) is slower than (7). This is
readily explainable by the temporary anomaly found in (8) and not in (7)
– *learning a drain* is strange but *learning a language* is not. Crucially, no
RT-difference was found at the embedded verb, suggesting that reanalysis
occurred in both conditions.

This matrix verb compatibility effect has also been observed in another
experiment where sentences such as (9) and (10) were compared in an ac-
ceptability study (Christensen, Kizach & Nyvad 2013).

(9) **Danish**

Hvilken båd foreslog naboen at vi skulle sælge

*Rather cheap*

Which boat suggested neighbor-the that we should sell

‘Which boat did the neighbor suggest that we sell rather cheap?’

(10) **Danish**

Hvor billigt foreslog naboen at vi skulle sælge

How cheap suggested neighbor-the that we should sell

‘How cheap did the neighbor suggest that we sell our boat?’

In (9) the fronted object, *hvilken båd* ‘which boat’, is a possible object
for the matrix verb *foreslog* ‘suggested’, but in (10) the fronted adverbial,
*hvor billigt* ‘how cheap’, is not a plausible modifier of the matrix verb. The
result showed that (9) was judged to be significantly more acceptable than
(10), which suggests that the temporary anomaly during processing has a
negative effect on acceptability.

By now it should be clear that processing difficulty – measured as a
slowdown in RT – typically leads to reduced acceptability. The question to
be investigated in this article is whether a slowdown that is not caused by a
temporary anomaly or reanalysis also lowers acceptability.

To test this, the following contrast were examined:
The only difference between (11) and (12) is the matrix verb – the verb *bemærkede* ‘noticed’ can take either a nominal or a sentential complement ([DP/CP]), whereas the verb *formodede* ‘presumed’ can only take sentential complements ([CP-ONLY]). In a self-paced reading study, Kizach et al. (2013: 7–8) demonstrated that there is reanalysis at the embedded verb only in (11). In other words, the subject of the embedded clause, *grisen i stalden* ‘the pig in the pen’, is temporarily misanalysed as an object in (11), whereas there is no such temporary misanalysis in (12).

![Figure 1: Mean RT, word-for-word, from Kizach, Nyvad & Christensen (2013: 8; experiment 3). Error bars ± 1 SE, *p<0.05, **p<0.001.](image)
The 1654 ms slowdown due to reanalysis at the embedded verb, \textit{manglede} ‘needed’, is quite clear (see Figure 1), but notice that there is a large 1583 ms slowdown at the word \textit{grisen} ‘pig-the’ in the \textit{CP-ONLY} condition. Kizach, Nyvad & Christensen (2013: 8–9) argued that this slowdown reflects the increased processing load needed to construct not only the DP \textit{grisen} ‘pig-the’ itself, but also the embedded clause, and since the verb \textit{formode} ‘presumed’ can only take CP-complements, the parser has no choice but to ‘spend time’ adding this structure to the parse tree. The interesting thing is that this processing difficulty is not due to some temporary misanalysis or reanalysis – it is simply an unavoidable hurdle. In both conditions a slowdown in RT can be measured, and the question is if both of these will decrease acceptability.

2. The acceptability judgment experiment

The idea was to see if the slowdowns we see in Figure 1 have any effect on acceptability. It is conceivable that they both have the same effect, in which case no acceptability difference is expected. Alternatively only one of them has an adverse effect, in which case there will be an acceptability difference between the two conditions, repeated below in (13) and (14) for convenience.

\begin{enumerate}
\item[(13)] \textbf{Danish – \textit{CP/DP} condition}
\begin{itemize}
\item Mia bemærkede grisen i stalden manglede vand.
\item \textit{Mia noticed pig-the in pen-the needed water}
\item ‘Mia noticed the pig in the pen needed water.’
\end{itemize}

\item[(14)] \textbf{Danish – \textit{CP-ONLY} condition}
\begin{itemize}
\item Mia formode grisen i stalden manglede vand.
\item \textit{Mia presumed pig-the in pen-the needed water}
\item ‘Mia presumed the pig in the pen needed water.’
\end{itemize}
\end{enumerate}

Based on the previous research the condition where the slowdown is due to reanalysis, i.e. (13), is predicted to have a reduced acceptability, but whether or not the unavoidable slowdown in (14) also reduces acceptability is unknown.

2.1 Participants, Materials and Methods

The 10 sentence pairs used in Kizach, Nyvad & Christensen (2013) were used in the experiment, because the word-for-word mean RT is known for...
these stimuli (see Figure 1). The only difference between the conditions were the matrix verb, and to control for lexical frequency effects, the token frequency of each verb was calculated using the Danish on-line corpus KorpusDK, available at https://ordnet.dk/korpusdk, and the two sets of frequencies were compared using the TOST-test for equivalence (Juzek 2016; chapter 4). Both t-tests were significant ($t_1(18) = -2.45, p \leq 0.05$; $t_2(18) = 4.02, p \leq 0.05$) suggesting that the two samples are equivalent – in other words, the frequencies of the matrix verbs in the two conditions are comparable.

The 10 experimental sentence pairs were divided into two lists, so that no participant saw the same item in both conditions, and 15 fillers were added to each list. The fillers ranged from completely acceptable (15) to completely unacceptable (16) sentences. Google Forms on Google Drive was used to create the lists and collect the data.

(15) **Danish**

Sonja talte i telefon med en veninde.

*Sonya spoke in phone with a friend*

’Sonya talked on the phone with a friend.’

(16) **Danish**

*Omend ham så gik det jo alligevel.

*Although him so went it nevertheless anyway*

‘Even though him it went ok nevertheless anyway.’

Links to the lists were made available on-line on the Facebook site Psycholab (a forum for students at Aarhus University interested in syntax) and sixty-four people participated (21 males). The mean age was 24.6 years with a range from 20 to 48.

The following instructions (in Danish, but here translated into English) were shown at the beginning of each list: “Judge the sentences on a scale from 1 (completely unacceptable) to 7 (completely acceptable). Try to follow your immediate intuition, and do not be affected by what you have been taught in school – there are no right or wrong answers here.”

### 2.2 Results

The CP-ONLY condition was judged to be more acceptable than the CP/DP condition (see Table 1).
### Table 1. Mean acceptability

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CP/DP (notice)</strong></td>
<td>4.51</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>CP-ONLY (presume)</strong></td>
<td>5.16</td>
<td>0.10</td>
</tr>
</tbody>
</table>

To test if the observed difference was statistically significant, the data was analyzed with a linear mixed-effects model following common practice in the field (Gibson, Piantadosi & Fedorenko 2011; Sprouse 2008). To perform the analysis the software R and the R-package `lmerTest` were used (Kuznetsova, Brockhoff & Christensen 2015; R Development Core Team 2015).

The acceptability score was the dependent variable and condition was the independent variable (neither gender nor age had any effect, and both variables were excluded from the final model reported here). The so-called maximal model was fitted to the data (Barr et al. 2013), and since comparisons with the zero-correlation-parameter model did not justify a simpler model (Bates et al. 2015) the maximal model is reported.

The result showed that the CP-ONLY condition was more acceptable than the CP/DP condition (see Table 2). The modest difference between the means (0.65) was indeed significant (p=0.047).

### Table 2. Summary of results of the linear mixed-effects model.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.162</td>
<td>0.317</td>
<td>16.272</td>
<td>0.000</td>
</tr>
<tr>
<td>CP/DP</td>
<td>-0.650</td>
<td>0.284</td>
<td>-2.291</td>
<td>0.047</td>
</tr>
</tbody>
</table>

### 2.3 Discussion

The result shows that reanalysis is costly as predicted – the CP/DP condition is consistently judged to be less acceptable than the CP-ONLY condition despite the fact that both types of sentences are perfectly grammatical in Danish. It makes sense that when the parser has to alter an established structure (changing the pig in the pen from being the direct object of observed into being the subject of needed) a price has to be paid, so to speak. This additional processing effort can be measured both in RT (as shown in Figure 1) and in acceptability. It is interesting that the reanalysis from direct object to subject after verbs such as observe has been discussed as cost free reanalysis in earlier literature on parsing (Fodor & Inoue 1998: 120; Pritchett 1992: 109–110), but evidently even this low cost can be measured.
It is tempting to interpret the result as if the slowdown in the \textsc{cp-only} condition is cost free, but logically it is only possible to conclude that the cost for this slowdown is less than for the one caused by reanalysis. Ole Togeby (personal communication) has pointed out that the RT contrast is greater for reanalysis (398 ms) than for the unavoidable structure building (259 ms), so it is possible that the acceptability difference reflects this difference. The result seems to suggest that the relation between a slowdown in RT and acceptability is not completely linear: A reanalysis slowdown is at least more costly than an inevitable slowdown. Building a representation of an embedded clause requires more structure than simply adding a direct object, and the increase in processing load in this situation may not be cost free, but at least it is less costly than reanalysis – a difference that can be detected in acceptability, but not in self-paced reading where both induce a significant slowdown (1583 ms and 1654 ms respectively).

3. Conclusion
As it turns out, all slowdowns are actually not created equal. Reanalysis affects RT as well as acceptability. Slowdown caused by inevitable syntactic structure building has a smaller effect on acceptability, or perhaps no effect. Hawkins (2004: 51–52, 155–156) suggests that the addition of a complementizer will greatly improve the processing ease of sentences like (14), and demonstrates in a corpus study that the omission of the complementizer \textit{that} in English is correlated with the length of the embedded subject – i.e. if the subject is a nominative pronoun, \textit{that}-omission is the norm, but with DP-subjects longer than 2 words \textit{that}-omission is much rarer. If the constraints on complementizer omission are (at least) similar in Danish, it would be possible to test if the slowdown in the \textsc{cp-only} condition affects acceptability or not. First it would have to be established that the slowdown observed in the \textsc{cp-only} condition at the embedded subject disappears when a complementizer is added. If this is the case, then one could compare \textsc{cp-only} conditions with and without a complementizer in an acceptability judgment experiment, and any difference would then reflect the cost of the slowdown. Perhaps someone decides to test this in the future. For now it is at least certain that inevitable structure building and reanalysis both increase RT, but measured in terms of acceptability, reanalysis is more costly.
References


