Extraction from RCs in Swedish:
What can we learn from spontaneously produced examples?

Symposium on syntactic islands in Scandinavian and English,
June 11–12, 2019, Aarhus University
Filippa Lindahl
Outline

- Introduction
- What can we learn from production data?
- What kind of extraction domains are Swedish RCs?
- Concluding remarks
Introduction

• Sentences like (1) (henceforth ERC-sentences) have been discussed in the mainland Scandinavian grammatical literature at least since the 1890s.

(1) Det₁ er der ingen [der har sagt mig __₁].
    that is there no one who have said me
    ‘No one has told me this.’
    (Mikkelsen 1894, p. 441)

(2) [Mot myggor]₁ finns det folk [som använder eukalyptusolja __₁].
    against mosquitos exist there people that use eucalyptus oil
    ‘There are people who use eucalyptus oil against mosquitos.’
    (Wellander 1937, p. 507)
In the 1970s and 80s these sentences came to the attention of international syntax research (Erteschik-Shir 1973, Allwood 1976, Andersson 1982, Christensen 1982).

Several Scandinavian linguists pointed out that mainland Scandinavian languages were exceptions to constraints like Ross's (1967) CNPC.

A few of many attempts to pin down the conditions on ERC:

- ERC is context sensitive, information structure is important (Erteschik-Shir 1973, Engdahl 1997)
- Certain predicates are more natural sounding in ERC-sentences (Erteschik-Shir 1973, Engdahl 1997)

Reference grammars describe ERC as acceptable in these languages (given that certain conditions are fulfilled) (Faarlund et al. 1997, Teleman et al. 1999, Hansen & Heltoft 2011).
A few recent studies


• It has been suggested both that RCs are non-/weak islands in these languages, and that they are strong islands and seemingly acceptable ERC is only illusory.

• Many of the recent studies are based on experimental data from formal acceptability judgement tasks.

• It is interesting to note that the overall acceptability of ERC is often reported to be quite low.
Low acceptability

• Christensen & Nyvad (2014) and Müller (2015) set out to test the small clause hypothesis, i.e. that escapable islands in Scandinavian are SC complements (Kush 2011, Kush, Omaki & Hornstein 2013).

• While neither found support for the hypothesis, Christensen & Nyvad (2014) found that acceptability correlated with lexical frequency of the matrix predicate.

• In both of the studies, acceptability of ERC was quite low with most predicates.
  – Christensen & Nyvad (2014): ERC-sentences with all verbs except three (møde ‘meet’, se ‘see’, kende ‘know’) cluster between 1 and 3,5 on a seven-point Likert scale.
  – Müller (2015): Acceptability of ERC between 2.96 and 3.32 on average on a seven-point Likert scale.
• Kush, Lohndal & Sprouse (2018):
  – Investigated *wh*-extraction from five island types.
  – Clear evidence of subject and adjunct island effects.
  – *Wh*-extraction from RCs was judged just as unacceptable as extraction from subjects and adjuncts.

• Kush Lohndal & Sprouse (to appear):
  – A similar study investigating topicalization out of the same island types showed a partially different pattern.
  – Extraction from subjects still highly unacceptable.
  – Inter-individual and inter-trial variation in judgements of extraction from embedded questions and RCs suggests that these two domains are not uniform syntactic islands.
Questions

• How should we interpret the fact that ERC-sentences often get low acceptability scores in formal judgement tasks? Is ERC actually unacceptable/ungrammatical in mainland Scandinavian?

• What kind of domain do RCs constitute?
  – Strong islands?
  – Non-islands?
  – Weak islands?
Today’s talk

• In this talk, I aim to do two things:
  – Show ERC production data, and discuss how such data can inform our understanding of ERC.
  – Argue that RCs are weak islands in Swedish. (This argument is not solely based on production data, but also on a questionnaire and elicitation).
What can we learn from production data?
Data collection

• 270 spontaneously produced ERC-sentences from spoken and written Swedish (gathered between 2011 and 2017).
  – Spoken language data: From the radio, TV, and conversation (161 examples).
  – Written examples: From newspaper text, social media, and fiction (109 examples).
• I came across most of the examples in everyday life, but several of them were collected as an effort to find rare types of ERC, for example extractions involving a certain embedding predicate or type of extracted phrase.
• The data can be accessed at https://svn.spraakbanken.gu.se/sb-arkiv/pub/lindahl/2017/.
Examples from the wild

Conversation

mitt privata mobilnummer (. ) jag har ett 076  
My private cell number. I have a 076  
‘My private cell number. I have a 076 number’

det har jag aldrig träffat på nån annan [ som har ___ ]  
that have I never met on someone else that has  
‘I’ve never met anyone else who has that.’

det högg jag på  
That bit I on  
‘I jumped on that.’

(Conversation, 2016)
Stand-up performance

alkoholism är däremot ingen sjukdom. Det stör jag mej på folk [som säger __].

‘Alcoholism is however no disease. That annoy me on people that say.’

det är ingen sjukdom alkoholism. Det är bara skitsnack. Det är en hobby (LAUGHTER)

‘It is not a disease, alcoholism, that’s just bullshit. It’s a hobby.’

det är två olika saker

that is two different things.

‘That’s two different things.’

(Magnus Betnér, Folkhälsan, stand-up show, 2010)
Radio

gillar du inte åkertistel? det finns det många andra [ som gör ___ ]
like you not field thistle? that exist there many others that do
‘Don’t you like field thistles? There are many others who do.’

massor med fjärilslarver har åkertisteln som värdväxt
masses with caterpillars have field thistle as food plant
‘Many caterpillars have the field thistle as their food plant.’

(Swedish Radio, Odla med P1, July 9 2012)
Forum thread

Det sägs ofta om lipizzaner att de är enmanshästar och att de kan vara tuffa i psyket.
it says often about lipizzaner that they are one-man-horses and that they can be tough in mentality.
‘It’s often said about Lipizzans that they are one-man-horses and that they can be hardheaded.’

Det har jag inte träffat någon men det finns säkert undantag.
that have I not met any but there are surely exceptions
‘I’ve not met any that are, but I’m sure there are exceptions.’

(Forum thread 2008)
To speed up construction, the Moderate Party demands that the government makes simplifications in the rules. – It should be possible to have a shared a kitchen, for example.

There are many who’d want that, says Ewa Thalén Finné, Moderate Party spokesperson for housing.

(Dagens Nyheter, 13 Jan 2016)
Novel

Nej, det är klart att han inte ska. Anna-Britta tar mötet istället för honom. `No, of course he shouldn’t. Anna-Britta will go to the meeting instead of him.’

Men det är det nästan ingen [som vet om ___]. `But there is almost no one who knows that.’

(Jonas Sjöstedt, Sahara, p. 182)
Peer-reviewed journal article

[Att placeringen av objekt före icke-finita verb var en genuin del av det spontana talspråket under fornsvensk tid ]1 är det knappast någon forskare [ som ifrågasatt ___1 ].

`There is hardly any researcher who has questioned that the placement of objects before non-finite verbs was a genuine part of the spontaneous spoken language during the Old Swedish period.'

Production data – What is common?

Most common:

• Topicalization (topicalized phrase is usually pronominal).
• Subject relative (usually sentence final)
• Indefinite, quantified RC head. Often no overt noun.
• Main predicates: vara ‘be’, finnas ‘exist’, ha ‘have’, veta ‘know’, höra ‘hear’, se ‘see’

(3) Det₁ är det många [som gör __₁]
that is there many that do
‘There are many people who do that.’
# Matrix predicates

<table>
<thead>
<tr>
<th>Type</th>
<th>No of instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existential (finnas)</td>
<td>13</td>
</tr>
<tr>
<td>Presentational (vara)</td>
<td>61</td>
</tr>
<tr>
<td>Predicational copular (vara)</td>
<td>4</td>
</tr>
<tr>
<td>Other main verbs (ha, veta, se, höra etc)</td>
<td>13</td>
</tr>
<tr>
<td>Det-cleft (vara)</td>
<td>8</td>
</tr>
<tr>
<td>Reverse DEM-cleft (vara)</td>
<td>1</td>
</tr>
<tr>
<td>Sum</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: *Matrix predicate types in a sample of 100 ERC sentences from* (Lindahl 2017, p. 94)
Production data – A few crucial example types

Case connectivity

Is that so? Well, you don’t look tired or hungry.

Och dig₁/*du var det ingen [som tvingade iväg __₁ till kolchoskontoret and you.ACC/*you.NOM was there no one that forced away to the kolkhoz office

mitt i natten och dömde __₁ till tjugofem års straffarbete ].
in the middle of the night and sentenced to twenty-five years’ penal servitude

‘No one forced YOU to the kolkhoz office in the middle of the night and sentenced you to twenty-five years of penal servitude.’

(Ruta Sepetys, Strimmor av hopp, novel)
Binding reconstruction

Sin plånbok och kontouppgifter skulle man aldrig ge bort till någon på krogen, their REFLEX wallet and account info would one never give away to someone at pub.DEF

men [ sin kropp ]₁ är det många [ som ger bort ___₁ utan att tänka efter ].

but their REFLEX body is there many that give away without to think after

‘One would never give away one’s wallet to someone at the pub, but there are many people who would give away their body without thinking about it.’

(Svenska Dagbladet, 14 Nov. 2007)
Non-subject RCs

[just den här delen av matten]$_1$ kommer ni inte hitta nått jobb$_2$ [där$_2$ man behöver ___1___2],

precisely the here part of math.DEF will you not find some job where one needs

‘You won’t find any job where you need THIS part of math.’

(Conversation, June 2016)

It fits well, but it’s not very well ventilated. I easily get a bit sweaty on my back.

Fast det$_1$ har jag inte hittat någon ryggsäck$_2$ [Op$_2$ jag inte blir ___1 av ___2].

but that have I not found some backpack I not become of

‘But I haven’t found any backpack that I don’t get sweaty from.’

(Forum thread, June 2008)
Interim conclusions

• ERC occurs in written and spoken Swedish. It occurs both in spontaneous, informal settings, like conversations, unplanned speech, and blogs, and in edited texts, including radio newscasts, newspapers, novels, and peer-reviewed articles.

• The extraction sentences are unremarkable in context, and don’t cause any comment, correction, or disruption.

• A number of speakers produce these types of sentences, and no one seems to react adversely to them in spontaneous discourse.

• This suggests RCs are not strong islands in every Swedish speaker’s grammar.
Interim conclusions

• While the collection method employed here does not allow us to conclude anything too particular about frequencies of various features of ERC-sentences, there are some clear patterns.

• ERC-sentences are syntactically complex, but in spontaneously produced examples they are most often quite low in lexical content.

• Certain matrix verbs are much more common than others (presentational and existential contexts), subject relatives are most common, etc.

• An interesting point in itself is that many of these patterns seem to be tendencies, rather than absolutes. Interesting from the perspective of acquisition/learnability.
What kind of extraction domains are Swedish RCs?
Strong islands, non-islands, or weak islands?

• Given the previous Scandinavian research, facts from production data, and recent experimental studies, there seems to be a growing consensus that RCs are not strong islands in the grammars of all speakers of mainland Scandinavian languages.

• But what kind of domain do they constitute?

• In this section of the talk, I will present some data that suggest that Swedish RCs are a species of weak islands, and discuss a few distinctions that seem to matter for extractability.
Strong vs. weak islands

• Szabolcsi (2006):
  – Strong islands:
    No extraction is allowed (constructions with an appropriate resumptive pronoun may be allowed)
  – Weak islands:
    Some phrases can extract, others cannot

(Szabolcsi 2006, p. 480)

• Many types of WIs have been investigated, but a prototypical example is embedded questions. The dependency that is most often discussed is wh-question formation.
It’s been observed in previous research that ERC is most common with topicalization, and that topicalization from RCs is more acceptable than *wh*-question formation (Engdahl 1997, Teleman et al. 1999).

In my sample of 100 ERC-sentences, I had 93 examples with topicalization, and 7 examples with relativization (Lindahl 2017).

No *wh*-questions.

The recent studies by Kush et al. (2018, to appear) confirm that there is a clear difference in acceptability between *wh*-question formation and topicalization from RCs in formal judgement tasks.
Questionnaire + elicitation

• Elicitation on wh-question formation
  – The interview participants were graduate students of Scandinavian languages at the University of Gothenburg and all native speakers of Swedish.

• Exploratory questionnaire (When, if ever, can adjuncts be topicalized from RCs?):
  – Judgement task, 32 example sentences in context.
  – Participants judged whether the sentence was a natural follow up (naturlig), somewhat strange (lite konstig), or unnatural (onaturlig).
  – 16 native Swedish speakers between 24 and 79 years old, all of whom had finished high school at the time of participation.
Non-islands or weak islands?

- In order to show that RCs are not just plain non-islands in Swedish, following Szabolcsi’s definition, we need to show that there are some phrases that cannot extract from them.
- To be sure that extraction of this phrase is not just impossible for other reasons, it should be possible to extract the same type of phrase from a *that*-clause.
- In my elicitation, it was clear that there are phrases of just this kind. The clearest example is *varför ‘why’*. 
Varför ‘why’ and that-clauses:

(4) Varför tror du att Strindberg skrev Fröken Julie? ‘Why do you think that Strindberg wrote Miss Julie?’

a. Därför att ordvalet och stilen liknar Strindbergs övriga verk. ‘Because the choice of words and the style resembles Strindberg’s other works.’

b. För att chockera sin samtid. ‘In order to shock his contemporaries.’
(5) Varför känner du många [som har skrivit böcker]?
‘Why do you know many people who have written books?’

a. För att jag är med i Författarförbundets styrelse.
‘Because I am on the board of the Swedish Writer’s Union.’

b. För att chockera sin samtid.
‘In order to shock their contemporaries.’

• The fact that varför cannot be extracted demonstrates that RCs are not non-islands.
D-linking

• If RCs are weak islands, we would expect D-linking to improve extraction (Comorowski 1989, Cinque 1990).

(6) [Vilka språk]₁ är det många [ som talar ___₁ i Sverige ]?
which languages are there many that speak in Sweden
‘Which languages are such that there are many people who speak them in Sweden?’

(7) [Vilka språk]₁ känner du nån [ som talar ___₁ ]?
which languages know.rel you someone that speaks
‘Which languages are such that you know someone who speaks them?’
(8) Vad såg du någon [som förgiftade ___1]?

what saw you someone that poisoned

‘What is such that you saw someone who poisoned it?’
Weak islands and adjunct types

(9) a. * How did John ask <which problem to phrase __>?
   ‘What is the manner such that John asked which problem to phrase in that manner?’

b. * Why did John ask < whether to fire him __ >?
   ‘What is the reason such that John asked whether it is a good reason for firing him?’

c. ?? When did John ask < whether to fire him __ >?
   ‘What is the time such that John asked whether it is a good time for firing him?’

d. ? Where did John ask < whether to put/read the book __ >?
   ‘What is the location such that John asked whether to put this book there/- whether to read this book there?’

(Szabolcsi 2006:494–495, see also Rizzi 1990)
• I found a somewhat similar pattern in extraction from RCs.

(10) a. Var$_1$ är det många [som brukar jobba övertid ___1]?
where are there many that tend work overtime
‘Which place is such that there are many people who tend to work overtime there?’

b. ?När$_1$ vet du många [som brukar vara trötta ___1]?
when know.COGR you many that tend be tired
‘Which time is such that you know of many people who tend to be tired then?’

c. ??Hur$_1$ vet du många [som fick sitt första jobb ___1]?
how know.COGR you many that got their.REFLX first job
‘Which way is such that you know of many people who got their first job that way?’
Topicalization of adjuncts

• While *wh*-adjuncts are hard to extract, the questionnaire showed that many speakers accept topicalization of adjuncts of various kinds.
(29) Linda: Min syster och jag är oense om när vi ska äta ikväll. Hon tycker att vi ska äta klockan nio, men jag tycker att det är aldeles för sent.

‘Linda: My sister and I disagree about when to have dinner tonight. She thinks we should eat at nine, but I think that is way to late.’

a. Ida: Ja, [så sent]_1 vet jag ingen [som brukar äta middag t_1].

‘Ida: Yes, I don’t know of anyone who eats dinner that late.’

(Questionnaire 11, 5, 0)

b. Ida: Ja, jag vet ingen [som brukar äta middag så sent].

‘Ida: Yes, I don’t know of anyone who eats dinner that late.’

(Questionnaire 12, 2, 1)
(30) Anna: Mina barnbarn sjöng så fint att alla i kyrkan grät.
   Anna: my children sang so nicely that everyone in church.\text{def cried}
Individual variation

- Both the questionnaire study and in particular the elicitation indicate that there is some individual variation in what types of extractions speakers are willing to accept.
- E.g. in the elicitation, while all four participants could answer D-linked argument-\textit{wh}-questions, other types were harder for some speakers.
- Kush et al. (2018, to appear) have begun a systematic investigation of individual variation (both inter speaker and intra-speaker). This seems like a promising way forward.
Concluding remarks

• Production data has an important role to play in theories of extraction:
  – In the case of Swedish ERCs, we saw that while such sentences frequently receive low acceptability judgements in formal judgement task, they occur in spontaneous Swedish in various registers, and pass as entirely unremarkable in context.
  – Without production data, we might have though that ERC was just not acceptable in Swedish.

• Production data can also help creating better stimuli for elicitation and experimental work.
Concluding remarks

• In the last section of the talk, I showed that RCs are weak islands in Swedish, given Szabolcsi’s definition.

• Based on elicitation with four speakers, I could show that there are certain patterns with respect to *wh*-question formation out of RCs, that seem to align with patterns otherwise found in extraction from embedded questions.

• More systematic work is needed to see to what extent these patterns hold up for more speakers, and more generally, how ERC compares to extraction from other more well know weak islands.
References

Allwood, Jens 1976. The complex NP constraint as a non-universal rule and some semantic factors influencing the acceptability of Swedish sentences which violate the CNPC. In: Stillings, Justine (ed.), UMOP II. Amherst, MA: University of Massachusetts, pp. 1–20.


Cederschiöld, Gustaf 1897. Om svenska som skriftspråk [Swedish as a written language]. Gothenburg: Wettergren & Kerber.


Korp: [http://spraakbanken.gu.se/korp/](http://spraakbanken.gu.se/korp/)


