

ISLAND EXTRACTIONS IN THE WILD

A corpus study of adjunct and relative clause extractions in Danish and English

Christiane Müller & Clara Ulrich Eggers

MINDS, October 29



LONG-DISTANCE DEPENDENCIES

- Languages can have *long-distance dependencies* (see AS-08, pp. 8-9):

(1) [Which book]_i did Harry think [_i that Bill bought _i]?


- Other types of long-distance dependencies are *topicalization* (2) and *relativization* (3):

(2) [That book]_i Harry thought [_i that Bill should buy _i].

(3) This is [the book]_i that Harry thought [_i that Bill should buy _i].

LONG-DISTANCE DEPENDENCIES

- Long-distance dependencies can in principle apply across an indefinite number of clauses:

- (4)
- a. What_i did Bill buy _i?
 - b. What_i did you force Bill to buy _i?
 - c. What_i did Harry say you had forced Bill to buy _i?
 - d. What_i was it obvious that Harry said you had forced Bill to buy _i?

- But certain syntactic environments block the formation of such long distance dependencies!



ISLANDS

- **Relative clauses:**

(5) *[Which book]_i did John meet [a child who read __i]?

- **Subjects:**

(6) *[Who]_i did [pictures of __i] annoy Bill?

- **Adjunct clauses:**

(7) *[Who]_i did John arrive [after Bill kissed __i]?

- Such domains are termed *islands* (Ross 1967)
- The same effect arises with other dependencies:
(8) *[This girl]_i John arrived [after Bill kissed __i].



ISLANDS

- A central question in linguistic theory has been what the source of such island effects is.
- I.e. why is extraction possible e.g. in (9a) but not in (9b) or (9c)?

- (9) a. Who_i did John say [that Bill kissed _i]?
b. *Who_i did John arrive [after Bill kissed _i]? (adjunct island)
c. *Who_i did John know [a girl that kissed _i]? (relative clause island)

- Many accounts assume that extraction from islands is ruled out by syntactic *island constraints*, and that these island constraints apply universally:
 - The **Complex NP Constraint** (Ross 1967): Extraction from complex NPs is disallowed.
 - The **Condition on Extraction Domains** (Huang 1982): Extraction is disallowed from domains that are not properly governed (roughly, non-complements).



ISLAND EXTRACTIONS IN MSC.

- However, the claim that island constraints are universal has been challenged by data from the Mainland Scandinavian (MSc.) languages:

(10) *Relative clause extraction (by topicalization)*

a. [De blommorna]_i känner jag [en man som säljer __i].
those flowers know I a man who sells

'I know a man who sells those flowers.' (Swedish; Allwood 1982: 24)

b. [Dette bildet]_i kjenner eg [den måleren som har måla __i].
that picture know I the painter who has painted

'I know the painter who has painted this picture.' (Norwegian; Faarlund et al. 1997: 1099)

c. [Suppe]_i kender jeg [mange der kan lide __i].
soup know I many who can like

'I know many people who like soup.' (Danish; Erteschik-Shir 1973: 67)



ISLAND EXTRACTIONS IN MSC.

(11) *Adjunct clause extraction (by topicalization)*

a. [Sportspegeln]_i somnar jag [när jag ser __i].

sports program.the fall.asleep I when I watch

‘I fall asleep when I watch the sports program.’ (Swedish; Anward 1982: 74)

b. [Den saka]_i ventar vi her [mens de ordnar __i].

this thing wait we here while they fix

‘We are waiting here while they fix this thing.’ (Norwegian; Faarlund 1992: 117)

c. [Den vase]_i får du ballade [hvis du taber __i].

this vase get you trouble if you drop

‘You are in trouble if you drop this vase.’ (Danish; Hansen & Heltoft 2011: 1814)

- Examples like these violate island constraints that are assumed to apply universally.
- So: Is there cross-linguistic variation in islands?



FORMAL VS. INFORMAL DATA

- The claims that such structures are acceptable in MSc. are based on *informal* judgments.
- However, *formal* studies of Scandinavian island extractions yield rather low ratings:
 - Swedish relative clause extractions scored ratings on the lower half of the scale in studies by Müller (2015), Wiklund et al. (2017), and Tutunjian et al. (2017)
 - Kush et al. (2018): extraction from relative clauses and from (conditional) adjunct clauses in Norwegian yielded acceptability scores at the bottom end of the scale
- What is actually possible in MSc.???



ACCEPTABLE EXTRACTION IN ENGLISH?

- Anecdotal evidence of acceptable extractions from finite adjunct clauses (12) and relative clauses (13) in English:

(12) a. This is [the watch]_i that I got upset [when I lost _i]. (Truswell 2011: 175)

b. a stranger, from [that remote and barbarian Isle]_i which the Imperial Roman shivered [when he named _i], paused. (Haegeman 2004: 70)

(13) a. Violence is [something]_i that there are [many Americans who condone _i].
(McCawley 1981: 108)

b. John is [the sort of guy]_i that I don't know [a lot of people who think well of _i]
(Culicover 1999: 230)

➔ How different are English and MSc. really?



A CORPUS STUDY

- We conduct an exploratory corpus study on:
 - adjunct clause extraction (Danish & English) + relative clause extraction (Danish & English)
- Why a corpus study?
 - We want to explore naturally produced examples – this will provide further insights into what is in fact possible in Danish and English
 - If we can find examples in the corpora, this would demonstrate that island extraction is not just a peripheral phenomenon restricted to isolated constructed examples
 - Most acceptability studies of island constructions aim to use test sentences that are modelled on naturally occurring data.
 - ➡ We can uncover potential patterns or trends among the found extractions.



METHOD

- Danish corpora: *KorpusDK (+ BySoc, Samtalegrammatik)*
- English corpora: *British National Corpus (BNC),
Corpus of Contemporary American English (COCA)*
- complemented by examples retrieved from a Google search
- We used a combination of different search strings that can target relative clause and adjunct clause extractions (if the corpus contains any).
 - Restriction: Our corpora are not annotated in a way that makes it possible to search for extraction constructions *directly*.
 - ➡ This means that we can't get quantitative data or compare frequencies across constructions or across languages.
 - Rather, our results can provide informal insights into what appears to be possible at all, or “common”, among the found extraction instances.



ADJUNCT CLAUSE EXTRACTION - FINDINGS

DANISH:

- > 30 instances of extraction from *hvis* ('if')- and *når* ('when')-adjunct clauses in Danish

- (14) a. især [den sidste]_i ville jeg blive sur, [hvis du ikke får __i].
b. fordi du værdsætter hundenes liv højt <3 og [det]_i bliver jeg glad [når jeg læser __i].
c. Det er [et nødvendigt mineral]_i, som vi ville dø, [hvis vi ikke fik __i].
d. Her er lige [et par billeder af en frederiksboger, en dansk varmblood og nogle fjordheste]_i som jeg ville blive glad [hvis I gad bedømme __i]

(from Google search)

- A search on KorpusDK yielded > 600 instances of extraction from *at*-clauses (not islands; e.g. “[Den hest]_i er jeg glad for [at du kan lide __i].”)
 - Extraction from (some) adjunct clauses appears to be possible, but less frequent than e.g. *at*-extraction.



ADJUNCT CLAUSE EXTRACTION - FINDINGS

ENGLISH:

- ≈ 50 instances of extraction from *if*- and *when*-adjunct clauses (**only** by relativization)
- (15) a. Many of the exercises are [ones]_i that I would be surprised [if even 1 percent of healthy women can do __i]. (COCA)
- b. Now, those are [things]_i that I feel very warm [when I look at __i], and I wouldn't want to live in a house that they - a house that didn't have room for those. (COCA)
- Naturally produced examples of adjunct clause extraction were found as easily in English as in Danish.
 - ➔ Potentially challenging the claim that English differs from the Scandinavian languages in never allowing extraction from finite adjunct clauses.



ADJUNCT CLAUSE EXTRACTION - FINDINGS

PATTERNS: ADJUNCT CLAUSE TYPE

- All examples found in English as well as in Danish featured *if*- and *when*-clauses, with extraction from *if*-clauses being the most common.
 - This result matches recent experimental findings that at least *if*-adjuncts do not behave like absolute islands for extraction in English (Nyvad et al. forthcoming).
- No examples of extraction from *because*-clauses were found.
 - ➔ Matching previous observations that conditional (*if/hvis*) and temporal (*when/når*) adjuncts appear to be more permissive for extraction than causal adjuncts (e.g. Müller 2019; Bondevik et al. 2020).



ADJUNCT CLAUSE EXTRACTION - FINDINGS

PATTERNS: EXTRACTION TYPE

- In Danish, most cases of adjunct clause extraction found involved relativization (and in English, **only** examples involving relativization were found).
 - Most existing acceptability studies of island structures have tested extraction in the form of topicalization and *wh*-movement.
 - ➡ Contrast between experimental designs and natural data



RELATIVE CLAUSE EXTRACTION

Preliminary findings – Danish:

- Examples of extraction from relative clauses almost always involve a presentational matrix verb (i.e. "presentational relatives"):

(16) a. [De store bøger]_i er det efterhånden næsten kun [humanister der skriver __i].

b. [Den gave]_i er vi [mange, der siger tak for __i]

c. Men [den slags spekulationer]_i kender jeg ikke [nogen, der tør fremsætte offentligt __i]

- These findings match the English examples of relative clause extraction reported to be acceptable in the literature (Chaves & Putnam 2020: 67).

Prediction for English:

- Examples will also primarily involve presentational matrix verbs.
- This would be a further indication that English and Danish are possibly more similar with regard to island extractions than previously assumed.



REFERENCES

- Allwood, Jens S. 1982. The complex NP constraint in Swedish. In Elisabeth Engdahl & Eva Ejerhed (eds.), *Readings on unbounded dependencies in Scandinavian languages*, 15–32. Stockholm: Almqvist & Wiksell.
- Bondevik, I., Kush, D., & Lohndal, T. 2020. “Variation in adjunct islands: The case of Norwegian”. *Nordic Journal of Linguistics*, December, 1–32.
- Chaves, R. P. & Putnam, M.T. 2020. *Unbounded Dependency Constructions*. Oxford: Oxford University Press.
- Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge, MA: MIT dissertation.
- Faarlund, Jan Terje. 1992. *Norsk syntaks i funksjonelt perspektiv* [Norwegian syntax in a functional perspective]. Oslo: Universitetsforlaget.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk* [Norwegian reference grammar]. Oslo: Universitetsforlaget.
- Haegeman, Liliane. 2004. The syntax of adverbial clauses and its consequences for topicalisation. In Martine Coene, Gretel De Cuyper & Yves D'hulst (eds.), *Current Studies in Comparative Romance Linguistics*, 61–90. Antwerp: University of Antwerp.
- Hansen, Erik & Lars Heltoft. 2011. *Grammatik over det danske sprog* [Danish reference grammar], vol. 3. København: Det danske Sprog- og Litteraturselskab: Syddansk universitetsforlag (kommissionær).
- Huang, James C.-T. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: MIT dissertation.



REFERENCES

- Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects: A case study of Norwegian Wh-extraction. *Natural Language & Linguistic Theory* 36(3). 743–779.
- Müller, Christiane. 2015. Against the Small Clause Hypothesis: Evidence from Swedish relative clause extractions. *Nordic Journal of Linguistics* 38(1). 67–92.
- Müller, C. 2019. *Permeable Islands: A Contrastive Study of Swedish and English Adjunct Clause Extraction*. PhD dissertation, University of Lund.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT Ph.D. thesis.
- Truswell, Robert. 2011. *Events, phrases and questions*. Oxford: Oxford University Press
- Tutunjian, Damon, Fredrik Heinat, Eva Klingvall & Anna-Lena Wiklund. 2017. Processing relative clause extractions in Swedish. *Frontiers in Psychology* 8:2118.
- Wiklund, Anna-Lena, Fredrik Heinat, Eva Klingvall & Damon Tutunjian. 2017. An acceptability study of long-distance extractions in Swedish. In Linda Escobar, Vincenç Torrens, Teresa Parodi (eds.), *Language processing and disorders*, 103–120. Newcastle upon Tyne: Cambridge Scholars Publishing.





AARHUS
UNIVERSITY

RELATIVE CLAUSE EXTRACTION

Extraction by topicalization – KorpusDK findings:

subject of the embedding clause Extracted element	<i>jeg/du/vi/I/de</i>	<i>han/hun/man</i>	<i>det</i>	<i>der</i>
“Heavy” element = headed by a noun ([ortho="(Den Det De)"][] {0, 2} [pos="N"] or [ortho="(Sådan Sådanne)"][] {0, 4} [pos="N"])	7 (1) [Den gave] _i , er vi [mange, der siger tak for ___].	1 (0) [Den lovovertrædelse] _i , er han næppe [den eneste, der har begået ___].	2 (0) [De store bøger] _i , er det efterhånden næsten kun [humanister der skriver ___].	36 (16) [De der æg] _i , er der da [ingen, der tager skade af ___].
Demonstrative/personal pronoun of the type “den/det/dem”, (En. “that/those”; [ortho="(Den Det Dem)"])	32 (30) [Det] _i , var vi [nogle stykker som var enige om ___].	3 (1) [Det] _i , var han [den eneste i familien, der kaldte mig ___].	2 (0) [Dem] _i , er det primært [centrum-venstre-partierne, der har svaret på ___].	442 (285) [Det] _i , er der [60.000 kunder, der gør ___ hver uge].
Personal pronoun of the type “mig/dig/hende/ham/os/jer”, (En. “me/you/her/ham/us/you”; [ortho="(Mig Dig Hende Ham Os Jer)"])	0 (0)	0 (0)	0 (1) ... men [ham] _i , var det nu [en fjollet studentskandale som havde fordrevet ___].	18 (4) [Mig] _i , er der overhovedet ikke [nogen, der skal bestemme over ___].

