

Constraints on wh-extraction in Czech

[preregistration]

anonymous

May 5, 2021

Contents

1 Overall design	3
1.1 Task, procedure, administration	3
1.2 Number of items	3
2 Experiment 1: Wh-extraction from adjuncts	4
2.1 Phenomenon and research question	4
2.2 Theories and hypotheses considered	4
2.2.1 Information structure theory	5
2.2.2 Phase- and anti-locality-based theory	5
2.2.3 Relativized minimality	6
2.3 Manipulated variables	7
2.4 Token set	8
2.5 Predictions	9
2.5.1 Information structure theory	9
2.5.2 Phases and anti-locality	10
2.5.3 Relativized minimality	11
2.6 Statistical analysis	11
3 Experiment 2: Left branch wh-extraction	11
3.1 Phenomenon and research questions	11
3.2 Theories and hypotheses considered	12
3.2.1 Müller’s analysis of Condition on Extraction Domain effects	12
3.2.2 Bondarenko and Davis’s analysis of scrambling and left branch extraction	13
3.3 Manipulated variables	14
3.4 Token set	14
3.5 Predictions	16
3.5.1 Müller’s analysis of Condition on Extraction Domain effects	16
3.5.2 Bondarenko & Davis’s analysis of scrambling and left branch extraction	16
3.6 Statistical analysis	17

4	Fillers	17
4.1	Filler batch 1	17
4.2	Filler batch 2	18
4.3	Filler batch 3	18
4.4	Filler batch 4	18
4.5	Filler batch 5	18
4.6	Filler batch 6	19

This document is a preregistration for two linguistic naturalness rating experiments. Section 1 characterizes the overall experimental design. Section 2 preregisters experiment 1, which deals with wh-extraction from clausal adjuncts in Czech. Section 3 preregisters experiment 2, which deals with word order-related restrictions on left-branch wh-extraction in Czech. Section 4 briefly characterizes the structure of the filler items.

1 Overall design

1.1 Task, procedure, administration

The task is to rate the naturalness (Czech: *přirozenost*) of Czech sentences on the scale 1 to 7, where 1 = completely unnatural (*zcela nepřirozená*) and 7 = completely natural (*zcela přirozená*). The stimuli are presented in written form using the software L-Rex (Starschenko & Wierzba 2021). The experiment is distributed to participants by email in the form of a URL link. The participants are friends, relatives, and acquaintances of the authors, mostly non-linguists, who do the experiment voluntarily in their homes. The pace is individual, but we expect the experiment to take between 20 and 30 minutes. The experiment will be administered between 5th May and 14th May and we aim to get between 50 and 100 participants.

The instructions, consent form, and all materials will be made available at the project page at OSF; see https://osf.io/zu9d4/?view_only=7000beb9b7f84fc3ae1a39fca355e4b0.

1.2 Number of items

The overall experiment consists of two subexperiments – experiment 1 and experiment 2 – and a set of fillers. The number of items per experiment/fillers is specified in Table 1. The subexperiments serve as mutual fillers, which means that there is $1.5\times$ more filler items than experimental items for experiment 1 and $2.5\times$ more filler items than experimental items for experiment 2. The total number of items (112) equals the number of stimuli that the participants will rate.

Experiment 1	48
Experiment 2	32
Fillers 1	8
Fillers 2	4
Fillers 3	4
Fillers 4	4
Fillers 5	6
Fillers 6	6
Fillers total	32
Total	112

Table 1: Number of items per experiment/fillers

2 Experiment 1: Wh-extraction from adjuncts

This experiment deals with the phenomenon of wh-extraction (aka wh-movement, wh-filler-gap dependency, etc.), illustrated in (1). The clause-initial wh-phrase *which book* is in a non-local dependency with the hypothetical gap (marked by *e*) in the complement of *read*, where the thematic role of the wh-phrase is encoded. Spelling out the metaphorical terminologies, we could say that the wh-phrase is “extracted” or “moved” from *e* or that wh-phrase “fills” *e*.

- (1) Which book did she read *e*?

Wh-extraction is subject to so-called locality restrictions. While extraction can in principle span long syntactic distances, as illustrated in (2a) or (2b), there are structures out of which a wh-phrase cannot be extracted. These structures are called syntactic islands (Ross 1967). Among the various island types are also the complex NP island (2c) and the adjunct island (2d).

- (2) a. Which book do you think [she read *e*]?
b. Which book do you think [she said [she read *e*]]?
c. *Which book did you consult [the teacher who discussed *e* in class]?
Intended meaning: ‘Which book was such that you consulted the teacher who discussed the book in class?’
d. *Which book did you consult the teacher [after he discussed *e* in class]?
Intended meaning: ‘Which book was such that you consulted the teacher after he discussed the book in class?’

2.1 Phenomenon and research question

The phenomenon investigated in experiment in 1 was first discussed by Lešnerová & Oliva (2003) and is illustrated in (3). What is of interest is the fact that the relative pronoun *kterou* ‘which’ appears to be extracted from the adjunct clause *když dítě odevzdá v ZOO* ‘when a child hands in [the sticker] in the zoo’.

- (3) Na každé zakoupené plyšové hračce je nálepka, kterou [když dítě odevzdá *e* v
on every bought stuffed toy is sticker which when child hands.in in
ZOO], obdrží navíc drobný dárek. (Lešnerová & Oliva 2003:241)
zoo gets in.addition small present
‘On every stuffed toy there’s a sticker such that if a child hands in the sticker in
the zoo, he/she will get a small present in addition.’

Our main research questions are:

- What makes the kind of extraction in (3) possible?
- What are the constraints on wh-extraction from clausal adjuncts in Czech?

2.2 Theories and hypotheses considered

In addressing the research questions, we consider (i) the information structure theory of syntactic locality (Erteschik-Shir 1973; Goldberg 2006; Abeillé et al. 2020) and two kinds

of syntactic theories: (ii) one based on syntactic phases and anti-locality (Bošković 2017; Biskup & Šimík 2019) and (iii) one based on relativized minimality (Rizzi 1990; Abels 2012). We discuss these briefly in turn.

2.2.1 Information structure theory

The core idea of the information structure theory of locality is that islands correspond to informationally backgrounded constituents, i.e., constituents that are topics, discourse-given, or are associated with a presupposition. The clausal complement of *think* is considered new information and therefore is not an island, (4a). The relative clause modifying the noun phrase in (4b), on the other hand, is likely to convey given or presupposed information and the noun phrase itself is definite and therefore associated with a presupposition. That makes the whole complex NP an island for extraction.

- (4) a. Which book do you think [she read *e*]?
 b. *Which book did you consult [the teacher who discussed *e* in class]?

According to a refined version of this theory (Abeillé et al. 2020), the extraction site and the extracted element must match in their information status. Interrogative wh-phrases stand for new or focused information, which is why they can be extracted from the new complement of *think* in (4a), but cannot be extracted from the given/presupposed complex NP in (4b).

This theory can handle Lešnerová & Oliva’s (2003) structure well because what is extracted is a relative pronoun, which is considered backgrounded by Abeillé et al. (2020), and it is extracted from an adjunct that is positioned to the left of its main clause, which plausibly maps to the adjunct’s backgrounded status. The example is repeated below for convenience.

- (5) Na každé zakoupené plyšové hračce je nálepka, kterou [když dítě odevzdá *e* v
 on every bought stuffed toy is sticker which when child hands.in in
 ZOO], obdrží navíc drobný dárek. (Lešnerová & Oliva 2003:241)
 zoo gets in.addition small present
 ‘On every stuffed toy there’s a sticker such that if a child hands in the sticker in
 the zoo, he/she will get a small present in addition.’

This theory further makes predictions with respect to (i) the position of the clausal adjunct and (ii) the kind of extracted element. We manipulate these variables in our experiment; see section 2.3. The specific predictions are specified in 3.5.

2.2.2 Phase- and anti-locality-based theory

According to this theory, long-distance wh-extraction proceeds via specified structural positions, called phase edges. There are two kinds of phases relevant for the present purposes: embedded clauses (CPs) and nominal phrases (NPs). In (6a), the wh-phrase *which book* is first extracted from e_1 to e_2 – the edge of the CP – and only then from e_2 to its final position. In (6b), an additional phase intervenes between e_2 and the final position of the wh-phrase; the wh-phrase would therefore have to move not just via e_2 , but also via e_3 . However, the step from e_2 to e_3 is too local and violates what is called anti-locality (Abels 2003).

- (6) a. Which book do you think [_{CP} e_2 she read e_1]?
 b. *Which book did you consult [_{NP} ! e_3 the teacher [_{CP} e_2 who discussed e_1 in class]]?

If we assume, following Biskup & Šimík (2019), that clausal adjuncts in Czech are simply CPs, this theory predicts a difference between (7a), which is a simplified version of Lešnerová & Oliva’s (2003) example, and (7b), where the complementizer *když* ‘when’ is replaced by *v případě, že* ‘in case that’, and the adjunct thus takes the form of a complex NP. While the CP adjunct is predicted to be transparent for extraction, the complex NP is predicted to be an island.

- (7) a. Je tam nálepka, kterou když dítě odevzdá, dostane dárek.
 is there sticker which when child hands.in gets present
 ‘There’s a sticker such that if a child hands the sticker in, he/she will get a present.’
 b. Je tam nálepka, kterou v případě, že dítě odevzdá, dostane dárek.
 is there sticker which in case that child hands.in gets present
 ‘There’s a sticker such that if a child hands the sticker in, he/she will get a present.’

In order to test this prediction, we manipulate the absence vs. presence of a nominal head of the clausal adjunct.

In addition, Biskup & Šimík (2019) argue that the position of the clausal adjunct relative to its main clause correlates with its syntactic status: left-peripheral adjuncts are considered to be CPs, playing the role of propositional restrictors of adverbial or modal quantifiers (Kratzer 1979, 2012), while right-peripheral adjuncts are considered to be free relatives, and hence NPs (Caponigro 2003), playing the role of temporal or other modifiers of the event expressed by the verb in a (Neo-)Davidsonian semantic representation (Davidson 1967; Parsons 1990). This assumption gives rise to the prediction that wh-extraction will be affected by the adjunct position: left-peripheral adjuncts are expected to be transparent for extraction, while right-peripheral ones, being NPs, are expected to be islands. This partly overlaps with the prediction of the information structure theory. Both are addressed by manipulating the clausal adjunct position in our experiment.

2.2.3 Relativized minimality

Relativized minimality is a theory of syntactic locality according to which only the structurally closest suitable constituent can be extracted, where structural closeness is “measured” from the landing site of the extracted element and is defined in terms of the c-command relation (Rizzi 1990). If a constituent A is to be extracted, but there is a closer and suitable constituent B, call it the intervener, A cannot be extracted, resulting in an island configuration. Here we subscribe to a version of relativized minimality where structural and semantic specificity (or “richness”) determines the suitability of the constituent (Starke 2001; Abels 2012). More particularly, B is an intervener for A if B contains all and possibly more syntactic and semantic features than A.

Consider again example (8) for illustration. We keep working with the standard assumption that wh-extraction proceeds in a successive cyclic fashion (cf. section 2.2.2). The wh-phrase *which book* in (8a) can move from e_1 to its intermediate landing site e_2 without any problems, since there is no intervener on the path between e_2 and e_1 . Sim-

ilarly, the *wh*-phrase can move to its final position in the matrix clause because nothing intervenes. In (8b), the *wh*-phrase cannot move from e_1 to e_2 because the relative pronoun *who* (or possibly the whole relative clause) is arguably more specific than the interrogative phrase *which book* and therefore, it acts as an intervener. Even if the *wh*-phrase were able to reach the edge of the relative clause, it would be inaccessible for extraction to its final position due to the NP, which arguably acts as an intervener.

- (8) a. Which book do you think [_{CP} e_2 she read e_1]?
 b. *Which book did you consult [_{NP} the teacher [_{CP} ! e_2 who discussed e_1 in class]]?

For the present purposes, we work with the specificity hierarchy in (9) (cf. Abels 2012 for a similar view): C corresponds to the declarative complementizer (or the embedded declarative clause), Wh corresponds to interrogative phrases (or the embedded interrogative clause) or more generally *wh*-like proposition-denoting embedded clauses such as conditionals (cf. Biskup & Šimík 2019), Rel corresponds to relative operators (or (free) relative clauses), and N corresponds to a nominal element selecting the relative clause, possibly a noun phrase or a determiner.

- (9) [N [Rel [Wh [C]]]]

We work with the hypothesis that Wh cannot move over Rel or N and Rel cannot move over N. But Rel *can* move over Wh. Furthermore, the movement of Wh across Wh leads to what is sometimes classified as a weak island.

This theory makes the same predictions as the phase- and anti-locality-based theory, but it makes one prediction in addition, namely that the extraction of a relative pronoun will be more natural than the extraction of a corresponding interrogative *wh*-phrase. Applied to our example, (10a) is expected to be more natural than (10b); (10b) is expected to be a weak island violation.

- (10) a. Je tam nálepka, kterou když dítě odevzdá, dostane dárek.
 is there sticker which when child hands.in gets present
 ‘There’s a sticker such that if a child hands the sticker in, he/she will get a present.’
 b. Zajímalo mě, kterou nálepku když dítě odevzdá, dostane dárek.
 interested me which sticker when child hands.in gets present
 ‘I wondered which sticker was such that when a child hands the sticker in, he/she will get a present.’

Manipulating the type of the extracted element (relative vs. interrogative) will thus be of use not only for testing the information structure theory, but also the relativized minimality theory.

2.3 Manipulated variables

We manipulated three mutually crossed variables, each with 2 levels, giving rise to 8 unique conditions ($2 \times 2 \times 2$), as summarized in Table 2.¹ All variables were manipulated within items and within subjects. ADJ-POSITION (= adjunct clause position) stands for the position of the adjunct clause relative to the main clause it modifies. The level left

¹Throughout, variables are typeset in SMALL CAPS and their levels in sans serif.

indicates that the adjunct clause is to the left of the main clause, as in *When it was raining, they stayed at home*, and the level **right** indicates that the adjunct clause is to the right of the main clause, as in *They stayed at home when it was raining*. The variable ADJ-HEAD (= adjunct clause head) stands for the absence or presence of a nominal that heads the adjunct clause. The nominal head is **absent** in cases like *They stayed at home when it was raining*. It is present in cases like *They stayed at home at the moment when it was raining*. The WH-TYPE variable stands for the type of the extracted wh-word: **rel** stands for the extraction of a relative wh-word, as in *They sold him a book which he wanted to read*, and **inter** stands for the extraction of an interrogative wh-phrase, as in *They asked which book he wanted to read*.

	ADJ-POSITION	ADJ-HEAD	WH-TYPE
a	left	absent	rel
b	right	absent	rel
c	left	present	rel
d	right	present	rel
e	left	absent	inter
f	right	absent	inter
g	left	present	inter
h	right	present	inter

Table 2: Manipulation of independent variables

2.4 Token set

The examples in (11) illustrate an item (item 10) in all of its 8 conditions. The subexample numbering corresponds to the a–h condition encoding in Table 2. What all examples have in common is a wh-extraction of *které (bonsaje)* ‘which (bonsais)’ out of the object position within the clausal adjunct – *když prodá* ‘when sells’ (level **absent**) or *v okamžiku, kdy prodá* ‘at the moment when sells’ (level **present**).

- (11) a. Pěstuje bonsaje, které když prodá, bude bohatý.
grows bonsais which when sells will.be rich
‘He grows bonsais which, when he sells them, he will be rich.’
- b. Pěstuje bonsaje, které v okamžiku, kdy prodá, bude bohatý.
grows bonsais which in moment when sells will.be rich
‘He grows bonsais which, at the moment that he sells them, he will be rich.’
- c. Pěstuje bonsaje, které bude bohatý, když prodá.
grows bonsais which will.be rich when sells
‘He grows bonsais which, when he sells them, he will be rich.’
- d. Pěstuje bonsaje, které bude bohatý, v okamžiku, kdy prodá.
grows bonsais which will.be rich in moment when sells
‘He grows bonsais which, at the moment that he sells them, he will be rich.’
- e. Ptal se, které bonsaje když prodá, bude bohatý.
asked REFL which bonsais when sells will.be rich
‘He asked which bonsais are such that when he sells them, he’ll be rich.’

- f. Ptal se, které bonsaje v okamžiku, kdy prodá, bude bohatý.
asked REFL which bonsais in moment when sells will.be rich
'He asked which bonsais are such that at the moment when he sells them, he'll be rich.'
- g. Ptal se, které bonsaje bude bohatý, když prodá.
asked REFL which bonsais will.be rich when sells
'He asked which bonsais are such that when he sells them, he'll be rich.'
- h. Ptal se, které bonsaje bude bohatý, v okamžiku, kdy prodá.
asked REFL which bonsais will.be rich in moment when sells
'He asked which bonsais are such that at the moment when he sells them, he'll be rich.'

The items for this experiment vary in a semi-systematic manner (between-items) in the following aspects:

1. the type of the adjunct clause: different types of temporal and conditional clauses (using complementizers/wh-operators such as *když*, *až*, *pokud*, *jakmile*, *kdykoliv*) and purpose clauses (complementizer *aby*);
2. the kind of the nominal element used (in conditions c, d, g, h): different kinds of nominals, preceded by suitable complementizers or wh-words and followed by suitable complementizers (*v případě*, *že* 'in case that', *za podmínky*, *že* 'on condition that', *v okamžiku*, *kdy* 'in moment when', etc.) or demonstratives, likewise preceded by prepositions and followed by complementizers (*potom*, *co* 'after.DEM COMP', *k tomu*, *aby* 'to DEM COMP', etc.);
3. the predicate embedding the interrogative clause (in conditions e–h): various kinds of predicates, e.g., 'know', 'remember', 'announce', 'say', 'ask', 'wonder'
4. the function of the extracted constituent within the adjunct clause: mostly object, sometimes adjunct, sometimes subject;
5. the animacy of the NP associated with the wh-phrase: human, animate, inanimate
6. the grammatical number of the NP associated with the wh-phrase: singular, plural
7. the presence of a pronominal (or *pro*) in the main clause coreferent with the wh-phrase/relative pronoun: mostly absent, sometimes present

2.5 Predictions

The predictions of the individual theories are specified below.

2.5.1 Information structure theory

The information structure theory predicts an interaction between ADJ-POSITION and WH-TYPE. Extraction of relative pronouns (*rel*), represented is expected to be more natural from left-positioned adjuncts than from right-positioned adjuncts and conversely for interrogative phrases (*inter*), where the extraction from right-positioned adjuncts is expected to be more natural than the extraction from left-positioned adjuncts. The reason for this is that backgrounded constituents (*rel*) should be relatively easily extractable from

backgrounded clauses (*left*), while focused constituents (*inter*) should be easily extractable from focused clauses (*right*).

It is not completely clear what the information structure theory predicts about the ADJ-HEAD factor. If the nominal head is **present**, though, the whole adjunct is probably treated backgrounded, independently of its position. Rel-extraction is thus expected to be more acceptable than *inter*-extraction, i.e., we might expect a main effect of the ADJ-HEAD factor.

The predictions are visually depicted in Figure 1.²

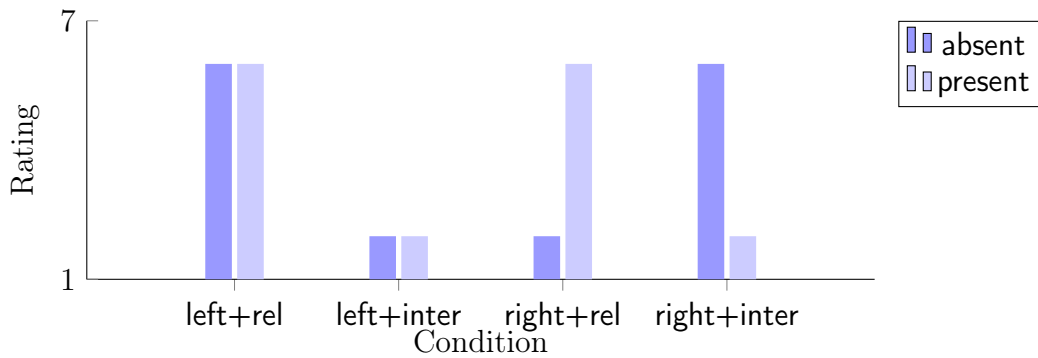


Figure 1: Exp 1: Prediction of information structure theory

2.5.2 Phases and anti-locality

The theory based on phases and anti-locality predict an interaction between ADJ-HEAD and ADJ-POSITION. The extraction is expected to be natural only in one of the four relevant conditions, namely out of *left* adjuncts with the nominal head **absent**. If the head is **present** (independently of the value of ADJ-POSITION), there is an NP phase to be crossed, leading to an anti-locality violation. If the adjunct is *right*, and even if the head is **absent**, the adjunct is represented by an NP, by hypothesis, leading to a violation of anti-locality. There is no specific prediction related to WH-TYPE.

The prediction is visualized in Figure 2.

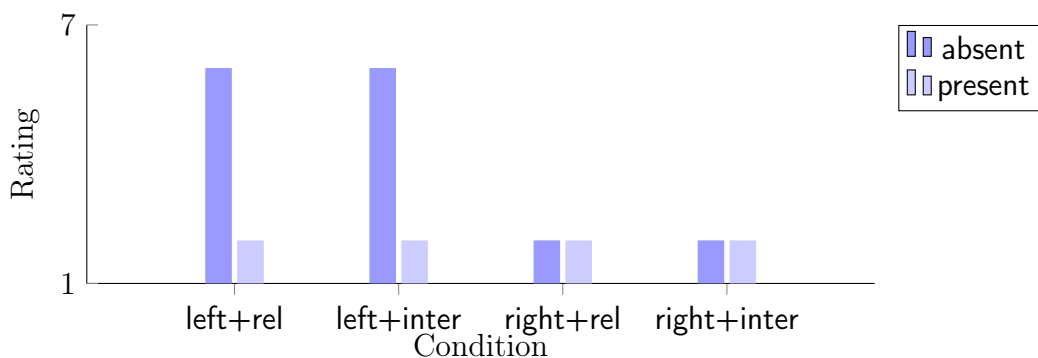


Figure 2: Exp 1: Prediction of phase- and anti-locality-based theory

²The values depicted represent expected mean ratings on the scale 1–7. The values are set to 6 and 2 merely for expository reasons. The prediction concerns the differences depending on the condition, not the absolute values.

2.5.3 Relativized minimality

The relativized minimality account predicts a three-way interaction between ADJ-POSITION, ADJ-HEAD, and WH-TYPE. Extracting a relative pronoun (*rel*) out of a left adjunct should be more natural than extracting an interrogative wh-phrase (*inter*) out of a left adjunct, which in turn should be more natural than any other condition. The reason for this is that the relative pronoun is syntactically and semantically specific enough to escape a proposition-denoting adjunct CP. The specificity of an interrogative wh-phrase, on the other hand, corresponds to the specificity of the adjunct complementizer; *inter*-extraction out of left adjuncts with an *absent* head are expected to have an effect comparable to the extraction out of weak islands. All other conditions involve an overt or a covert nominal projection, which are more specific than both *rel*- and *inter*-phrases.

The predicted pattern is visualized in Figure 3.

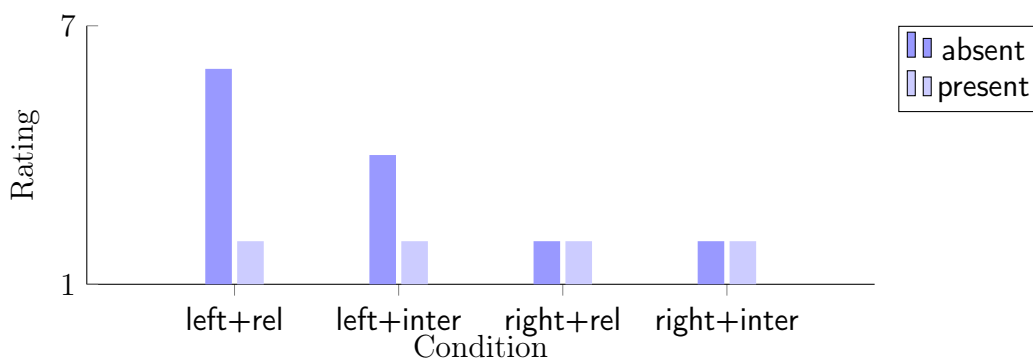


Figure 3: Exp 1: Prediction of phase- and anti-locality-based theory

2.6 Statistical analysis

In order to determine how the manipulated factors and their interactions affect the naturalness of sentences, we plan to fit linear mixed-effects models, using the *lmer* package (Bates et al. 2015) of R (R Core Team 2021), starting from the maximal random-effects specification justified by the design, following Barr et al. (2013).

3 Experiment 2: Left branch wh-extraction

This experiment focuses on wh-extraction from the left branch of a constituent of a transitive verb, as demonstrated in (12). The clause-initial wh-phrase *který* is in a dependency with the gap (marked by *e*) contained in the object of *korigovala*.

- (12) Který Jana korigovala [e článek]?
which Jana corrected article
'Which article did Jana correct?'

3.1 Phenomenon and research questions

The phenomenon of left branch extraction was first discussed by Ross (1967). The impossibility of displacement of a constituent on the left branch in languages like English was covered by his Left Branch Condition. The condition blocks displacement of the leftmost

nominal phrase from a larger nominal phrase (cf. also Emonds’s 1980 Generalized Left Branch Condition). As for Slavic languages, it has been observed that most of them allow left branch extraction; see e.g. Ross (1967) for Russian, Corver (1990) for Czech and Bošković (2005) for Serbo-Croatian.

Our main research questions are:

- Under which circumstances is left branch extraction of a wh-element like (8) possible in Czech?
- And conversely, which factors make left branch wh-extraction ungrammatical or less natural?

3.2 Theories and hypotheses considered

With respect to the research questions introduced above, we consider two recent syntactic theories discussing properties of movement and left branch extraction: Müller’s (2010) phase analysis of the Condition on Extraction Domain effects and Bondarenko & Davis’s (to appear) analysis of scrambling and left branch extraction from the subject in Russian. They are briefly discussed below.

3.2.1 Müller’s analysis of Condition on Extraction Domain effects

According to Huang’s (1982: 505) Condition on Extraction Domain, extraction from a phrase is possible only if the phrase is properly governed. This derives the asymmetry between complements and non-complements. Since complements – in contrast to specifiers and adjuncts – are properly governed, they allow extraction, as shown by the contrast in (13).

- (13) a. Who did you see [a picture of *e*]
 b. ?*Who does [a picture of *e*] hang on the wall? (Stepanov 2007:80)

Müller derives the Condition on Extraction Domain effects in a strictly derivational phase model in which all phrases are phases and intermediate movement steps are triggered by edge features. Since edge features can only be added before the phase head becomes inert, extraction from a phrase XP is blocked if the operation that merges XP is the final operation in a phase. In cases with extraction from the complement like (13a), an edge feature can be added before the specifier/last phrase is merged, hence movement is possible. In contrast, in cases of extraction from the subject like in (13b), the phase head becomes inert after merger of the subject and consequently no edge feature (which would trigger extraction from the subject) can be added.

Müller is also concerned with freezing effects, as illustrated by example (14), in which the topicalized PP is a barrier for further extraction. The effects are covered by his Freezing Generalization (stating that a trace may not be included in a moved phrase if the antecedent of the trace c-commands the moved phrase) and are derived by two specific restrictions on multiple edge feature insertion.

- (14) a. Who do you think that he will talk [to *e*]?
 b. *Who do you think that [to *e*] he will talk?

(Müller 2010:55; originally Postal 1972)

Müller argues that his analysis predicts the existence of melting effects. Specifically, a specifier of a phase that is normally an island for extraction (i.e. it is last-merged) should cease to be an island when some constituent becomes an outer specifier of the phase by movement. He illustrates this phenomenon with local scrambling in German and Czech; consider (15).

- (15) a. *Was haben [*e* für Bücher] [den Fritz] beeindruckt?
what have for books the Fritz impressed
'What kind of books impressed Fritz?'
b. Was haben [den Fritz]₂ [*e* für Bücher] t₂ beeindruckt?
what have the Fritz for books impressed
'What kind of books impressed Fritz?' (Müller 2010:61)

Example (15a) shows that the subject is a barrier for *was für* 'what for' split if it precedes the object *den Fritz*. The subject however allows extraction of *was* if the object undergoes local scrambling in front of the subject, as shown in (15b). Technically, since scrambling is feature-driven, the phase head *v* is active and can receive an edge feature for attracting an element out of the subject in the inner specifier before the object *den Fritz* is scrambled to the outer specifier position.

In order to test these predictions, experiment 2 manipulates (i) extraction out of subject vs. object and (ii) the relative position of subject and object.

3.2.2 Bondarenko and Davis's analysis of scrambling and left branch extraction

Bondarenko & Davis (to appear) focus on two types of movement in Russian, left branch extraction from subjects and object scrambling, and their interaction. Their analysis combines the phase model with a cyclic linearization theory of spellout. They show that left branch extraction is possible from both the subject and the object; consider example (16).

- (16) a. Èta včera [*e* devočka] pogladila kota.
this yesterday girl stroked cat
'This girl stroked the cat yesterday.'
b. Ètogo devočka pogladila [*e* kota].
this girl stroked kota
'The girl stroked this cat.' (Bondarenko & Davis to appear:(1))

With respect to extraction from the subject, as in (16a), they differ from the analysis of Müller (2010), who predicts the extraction to be ungrammatical (given that the subject is the last-merged element in the phase).

Bondarenko & Davis (to appear) also show that Russian allows scrambling of an object over the subject; consider (17a). Crucially, if object scrambling is combined with left branch extraction from the subject, the sentence becomes ungrammatical, as demonstrated in (17b).

- (17) a. Kota₁ èta devočka pogladila e₁.
 cat this girl stroked
 ‘This girl stroked the cat.’
- b. *Èta₂ kota₁ [e₂ devočka] pogladila e₁.
 this cat girl stroked
 ‘This girl stroked the cat.’ (Bondarenko & Davis to appear:(2))

Here, scrambling of the object has an opposite effect than proposed by Müller (2010). While in Müller’s analysis, scrambling of the object to the position in front of the subject melts the barrier of the subject, in Bondarenko & Davis’s (to appear) analysis, object scrambling to the edge of *v*P creates a *v*P linearization – with the object preceding the “whole” subject – contradictory to the linearization of the CP phase, with the left branch element extracted from the subject preceding the object. What is crucial in this respect is that the left branch cannot be extracted from the subject in the *v*P phase since there is a ban on movement from one specifier to another of the same phrase.

In order to test these predictions, experiment 2 manipulates (i) extraction out of subject vs. object and (ii) the relative position of subject and object.

3.3 Manipulated variables

We use three independent variables, each having 2 values, which results in 8 specific conditions ($2 \times 2 \times 2$), as shown in Table 3. The variables were manipulated within items and within subjects. **ADDITIONAL ARGUMENT** stands for the fact whether or not the second argument (from which no extraction happens) is present in the sentence. The value **covert** means that the second argument is only covertly present, as in *They were eating*, whereas the level **overt** indicates that the second argument is present overtly, as in *They were eating some fish*. The variable **EXTRACTION SITE POSITION** stands for the position of the argument from which left branch extraction happens. The value **final** means that the argument occurs in the clause final position, as in *Which did they eat [t fish]*, and the level **non-final** indicates that the appropriate argument occurs in a non-final position in the clause, as in *Which did [t workers] eat*. The **EXTRACTION SITE** variable stands for the type of the argument from which left branch extraction happens. The level **object** indicates that the argument from which an element is extracted is the object, as in *Which did they eat [t fish]*. The level **subject** means that it is the subject, from which an element is extracted, as in *Which did [t workers] eat*.

3.4 Token set

The following examples show an item in all of its 8 conditions. The numbering of subexamples in (18) corresponds to the a–h condition illustrated in Table 3. In all examples, a left branch extraction of an interrogative wh-element (*kterou* ‘which’ or *který* ‘which’) takes place in the embedded clause. The gap position can be reliably recovered by case-marking on both the wh-word and the corresponding noun. The verbs used in the items are all optionally transitive, allowing for object omission (relevant for conditions e and g). Subject omission is always grammatically possible in Czech, as Czech is a *pro*-drop language.

	ADDITIONAL ARGUMENT	EXTRACTION SITE POSITION	EXTRACTION SITE
a	covert	final	object
b	overt	final	object
c	covert	non-final	object
d	overt	non-final	object
e	covert	final	subject
f	overt	final	subject
g	covert	non-final	subject
h	overt	non-final	subject

Table 3: Manipulation of independent variables

- (18) a. Chtěli jsme vědět, kterou tady obědval polévku.
 wanted AUX.1PL know which.ACC here had.for.lunch soup.ACC
 ‘We wanted to know which soup he had for lunch.’
- b. Chtěli jsme vědět, kterou doktor obědval polévku.
 wanted AUX.1PL know which.ACC doctor.NOM had.for.lunch soup.ACC
 ‘We wanted to know which soup the doctor had for lunch.’
- c. Chtěli jsme vědět, kterou tady polévku obědval.
 wanted AUX.1PL know which.ACC here soup.ACC had.for.lunch
 ‘We wanted to know which soup he had for lunch.’
- d. Chtěli jsme vědět, kterou doktor polévku obědval.
 wanted AUX.1PL know which.ACC doctor.NOM soup.ACC had.for.lunch
 ‘We wanted to know which soup he had for lunch.’
- e. Chtěli jsme vědět, který tady obědval doktor.
 wanted AUX.1PL know which.NOM here had.lunch doctor.NOM
 ‘We wanted to know which doctor had lunch here.’
- f. Chtěli jsme vědět, který polévku obědval doktor.
 wanted AUX.1PL know which.NOM soup.ACC had.for.lunch doctor.NOM
 ‘We wanted to know which doctor had a soup for lunch.’
- g. Chtěli jsme vědět, který tady doktor obědval.
 wanted AUX.1PL know which.NOM here doctor.NOM had.lunch
 ‘We wanted to know which doctor had lunch here.’
- h. Chtěli jsme vědět, který polévku doktor obědval.
 wanted AUX.1PL know which.NOM soup.ACC doctor.NOM had.for.lunch
 ‘We wanted to know which doctor had a soup for lunch.’

The items for this experiment vary in a semi-systematic manner (between-items) in the following aspects:

1. the wh-word used: all extractions out of subjects use the wh-word *kteřý* ‘which’; extractions out of objects vary between *kteřý* ‘which’ and *jaký* ‘what.kind.of’
2. the animacy of the object: while the subject always refer to a human, some objects refer to humans, others are animate and yet others inanimate;
3. form of the additional argument: most additional arguments (subjects or objects) are realized as bare nouns, but some include a demonstrative determiner

3.5 Predictions

In this section, we specify predictions of the two syntactic theories discussed above.

3.5.1 Müller’s analysis of Condition on Extraction Domain effects

We consider Müller’s (2010) predictions separately for objects and subjects. For objects, we expect a main effect of EXTRACTION SITE POSITION: extraction from **final** objects should be more natural than extraction from **non-final** (scrambled) objects. This is because **non-final** objects are scrambled and in their scrambled positions they undergo freezing and hence become islands. For subjects, we expect an interaction between EXTRACTION SITE POSITION and ADDITIONAL ARGUMENT: extraction out of subjects should only be possible in case melting occurs, which is when the object scrambles over the subject to the edge of *v*P, thus ridding the subject off its island status. That happens only if the subject is **non-final** and the object **overt**. When the subject is **final** and the object is **overt**, intermediate movement of the object to the edge of *v*P takes place before the subject enters the derivation, leaving the subject with its island status. Given that the subsequent object scrambling targets a position outside of the *v*P, no melting effect is observable.

The predictions are visualized in Figure 4.

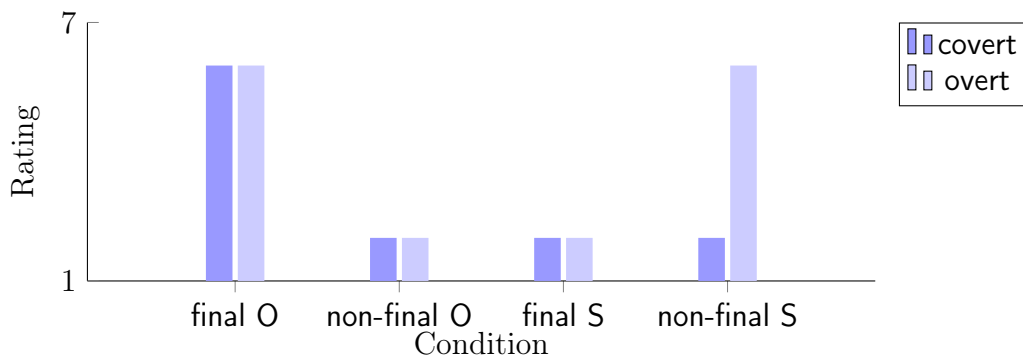


Figure 4: Exp 2: Prediction of Müller’s (2010) theory

3.5.2 Bondarenko & Davis’s analysis of scrambling and left branch extraction

Bondarenko & Davis (to appear) predict an interaction between EXTRACTION SITE POSITION and the overtiness of the ADDITIONAL ARGUMENT. In particular, the extraction out of **non-final** arguments is expected to be penalized, but only if it crosses an **overt** additional argument. It is in this configuration where the linearization of the arguments (and their subparts) at the level of *v*P is in conflict with the linearization at the level of CP. Bondarenko & Davis (to appear) do not incorporate a CED-like and freezing-like component that would render subjects or expressions in a derived position islands for extraction. Therefore, unless additional assumptions are made, no difference between extraction from **object** vs. **subject** is expected.

The prediction is visualized in Figure 5.

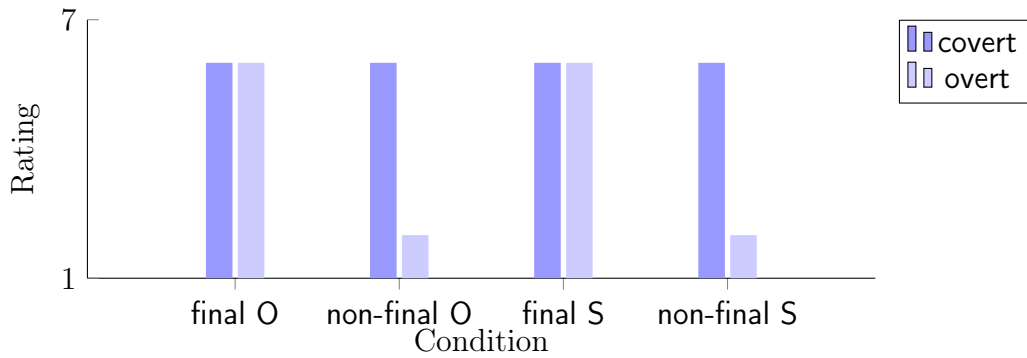


Figure 5: Exp 2: Prediction of Bondarenko & Davis’s (to appear) theory

3.6 Statistical analysis

In order to determine how the manipulated factors and their interactions affect the naturalness of sentences, we plan to fit linear mixed-effects models, using the lmer package (Bates et al. 2015) of R (R Core Team 2021), starting from the maximal random-effects specification justified by the design, following Barr et al. (2013).

4 Fillers

The fillers serve two purposes – to distract the participants’ attention from the critical manipulations and to control for some factors not directly manipulated in the designs of experiment 1 and experiment 2. There are 6 filler batches, each conceived of as a little subexperiment.

4.1 Filler batch 1

Fillers 1 manipulate different types of extractions; they include 8 conditions:

1. long interrogative extraction out of a complement clause;
2. long interrogative extraction out of a complement clause headed by a demonstrative (complex NP island);
3. short interrogative extraction;
4. complement clause without any extraction;
5. long relative extraction out of a complement clause;
6. long relative extraction out of a complement clause headed by a demonstrative (complex NP island);
7. short relative extraction;
8. relative clause with a complementizer (without a relative pronoun)

4.2 Filler batch 2

Fillers 2 manipulate the linear length of the filler-gap dependency; they include 4 conditions:

1. linearly short interrogative extraction (verb close to the extracted element);
2. linearly long interrogative extraction (verb far from the extracted element);
3. linearly short relative extraction;
4. linearly long relative extraction

4.3 Filler batch 3

Fillers 3 increase the structural length of the extraction in experiment 1 by one additional clause (not just *sticker which when...* but *sticker which I think that when...*) or by including a sentence adverbial. There are 4 conditions:

1. long relative extraction from left adjunct across one additional clause;
2. long relative extraction from right adjunct across one additional clause;
3. relative extraction from left adjunct across a sentence adverb;
4. relative extraction from right adjunct across a sentence adverb

4.4 Filler batch 4

Fillers 4 test the baseline naturalness of adjunct clauses depending on their position (left vs. right) and the presence of a nominal head; there are 4 conditions:

1. right-adjunct clause with a nominal head;
2. right-adjunct clause without a nominal head;
3. left-adjunct clause with a nominal head;
4. left-adjunct clause without a nominal head

4.5 Filler batch 5

Fillers 5 test long-distance extraction out of different kinds of clauses depending on the wh-type; there are 6 conditions:

1. relative extraction out of a declarative complement clause;
2. relative extraction out of a polar interrogative complement clause;
3. relative extraction out of a wh-interrogative complement clause;
4. interrogative extraction out of a declarative complement clause;
5. interrogative extraction out of a polar interrogative complement clause;
6. interrogative extraction out of a wh-interrogative complement clause;

4.6 Filler batch 6

Fillers 6 test different word orders in embedded clauses, serving as a control for experiment 2; there are 6 conditions:

1. SVO
2. SOV
3. OSV
4. OVS
5. SV
6. VS

References

- Abeillé, Anne, Barbara Hemforth, Elodie Winckel & Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204. 104293. <https://doi.org/10.1016/j.cognition.2020.104293>.
- Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding*. Storrs, CT: University of Connecticut dissertation. <https://opencommons.uconn.edu/dissertations/AAI3104085>.
- Abels, Klaus. 2012. The Italian left periphery: A view from locality. *Linguistic Inquiry* 43(2). 229–254. https://doi.org/10.1162/LING_a_00084.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Biskup, Petr & Radek Šimík. 2019. Structure of conditional and (cor)relative clauses: New evidence from locality. In Maggie Baird & Jonathan Pesetsky (eds.), *NELS 49: Proceedings of the 49th Annual Meeting of the North East Linguistic Society, Volume 1*, 135–144. Amherst, MA: GLSA Publications.
- Bondarenko, Tanya & Collin Davis. to appear. Sub-extraction asymmetries and linearization in Russian. In *Formal Approaches to Slavic Linguistics 28: The Second Stony Brook Meeting 2019*, Ann Arbor, MI: Michigan Slavic Publications. <https://ling.auf.net/lingbuzz/004915>.
- Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59(1). 1–45. <https://doi.org/10.1111/j.1467-9582.2005.00118.x>.

- Bošković, Željko. 2017. Extraction from complex NPs and detachment. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax: Second edition*, chap. 23. Wiley Blackwell. <https://doi.org/10.1002/9781118358733.wbsyncom023>.
- Caponigro, Ivano. 2003. *Free not to ask: On the semantics of free relatives and wh-words cross-linguistically*. Los Angeles, CA: University of California dissertation.
- Corver, Norbert. 1990. *The syntax of left branch extractions*. Tilburg: Tilburg University dissertation.
- Davidson, Donald. 1967. The logical form of action sentences. In Nicholas Rescher (ed.), *The logic of decision and action*, 81–95. Pittsburgh: University of Pittsburgh Press.
- Emonds, Joe. 1980. Word order in generative grammar. *Journal of Linguistic Research* 1. 33–54.
- Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge, MA: MIT dissertation. <http://dspace.mit.edu/handle/1721.1/7582>.
- Goldberg, Adele. 2006. *Constructions at work*. Oxford: Oxford University Press.
- Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammars*. Cambridge, MA: MIT dissertation. <http://dspace.mit.edu/handle/1721.1/15215>.
- Kratzer, Angelika. 1979. Conditional necessity and possibility. In Rainer Bäuerle (ed.), *Semantics from different points of view*, 117–147. Berlin: Springer.
- Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.
- Lešnerová, Šárka & Karel Oliva. 2003. Česká vztažná souvětí s nestandardní strukturou. *Slovo a slovesnost* 64(4). 241–252. <http://sas.ujc.cas.cz/archiv.php?art=4129>.
- Müller, Gereon. 2010. On deriving the CED from the PIC. *Linguistic Inquiry* 41(1). 35–82. <http://www.jstor.org/stable/40606828>.
- Parsons, Terence. 1990. *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.
- Postal, Paul M. 1972. On some rules that are not successive cyclic. *Linguistic Inquiry* 3(2). 211–222. <https://www.jstor.org/stable/4177702>.
- R Core Team. 2021. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.r-project.org>.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT dissertation. <http://hdl.handle.net/1721.1/15166>.
- Starke, Michal. 2001. *Move dissolves into merge: A theory of locality*. Geneva: University of Geneva dissertation. <http://ling.auf.net/lingbuzz/000002>.
- Starschenko, Alexej & Marta Wierzba. 2021. L-Rex Linguistic rating experiments [software], version beta. GNU General Public License v3.0. <https://github.com/2e2a/l-rex/>.

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1). 80–126. <https://doi.org/10.1111/j.1467-9612.2007.00094.x>.