Univerzita Karlova

Filozofická fakulta

Ústav českého jazyka a teorie komunikace

Bakalářská práce

Daniela Kořánová

Left branch extraction and basic word order in Czech

Extrakce levé větve a základní slovosled v češtině

Praha 2023

Vedoucí práce: doc. Mgr. Radek Šimík, Ph.D.

Acknowledgements

This thesis would not be possible without the excellent supervision of doc. Mgr. Radek Šimík, Ph.D. I am grateful for his comments and enthusiasm for the topic, he was always ready to help, even when technical difficulties came to the surface. Many thanks also go to my family and friends for their support and patience when they had to listen to my explanations and constant questions about the sentences I was creating for the experiment. Lastly, special thanks to doc. Mgr. Jan Chromý, Ph.D. for sending out the experiment to participants.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 10. dubna 2023

…..…….

Daniela Kořánová

Abstract

The topic of this bachelor's thesis is left branch extraction (LBE) in Czech. LBE is a non-projective construction created by the division of the nominal phrase, more precisely by the movement of the left constituent (e.g., Kterou$_i$ jsi viděl [t_i dívku]$_{NP}$?). The construction is limited, e.g., the left branch cannot be extracted from an NP that is complement of PP. More importantly, LBE is restricted by the word order. Using the analysis based on the phase theory and the theory of cyclic linearization (which proved to be sufficient), LBE is only possible if no scrambling of constituents happens (i.e., anti-locality problem does not occur). The two main factors that influence the word order are the syntactic function and the animacy (humanness) of the constituents. Based on the results of the conducted experiment, the more important factor for the ordering of words is the humanness. If this is taken out of the equation (i.e., the humanness of both subject and object is balanced), the word order is based on the syntactic function. For the possibility of the canonical word order to be based on the humanness, the analysis reflecting the uniformity of theta-assignment hypothesis (UTAH) and base-generating human object on a higher SpecvP to a nonhuman subject agrees with the results. For clause with LBE to be natural (acceptable), the word order needs to be a human subject preceding a nonhuman object. If the humanness is balanced, the word order of the clause has to be subject – object.

Key words: left branch extraction, word order, Czech syntax, phases, linearization, semantics

Abstrakt

Tématem této bakalářské práce je extrakce levé větve (LBE) v češtině. Jedná se o neprojektivní věty vzniklé rozdělením jmenné fráze, konkrétně posunem jejího levého konstituentu (např. Kterou$_i$ jsi viděl [t_i dívku]$_{NP}$?). Tato konstrukce je omezena různými faktory, např. extrahovaná levá větev nesmí být částí jmenné fráze, která je komplementem předložkové fráze. Práce se především zaměřuje na omezení přijatelnosti vět s LBE na základě slovosledu. Analýza, která se potvrdila výsledky provedeného experimentu, je založena na teorii fází a teorii cyklické linearizace. Podle této analýzy by měla být extrakce levé větve možná pouze za předpokladu, že ve větě nedochází ke scramblingu, tj. nenastane problém anti-lokality. Dva hlavní faktory ovlivňující slovosled jsou syntaktická funkce konstituentů a jejich životnost. Výsledky experimentu ukázaly, že druhý zmíněný faktor má přednost v uspořádání slovosledu. Pokud je životnost subjektu stejná jako životnost objektu (životnost je vyrovnaná), slovosled je na základě syntaktické funkce. Pro to, aby kanonický slovosled mohl být životný objekt následovaný neživotným subjektem, je třeba analýzy, která se řídí hypotézou o uniformitě připisování theta rolí (UTAH). Životný objekt je bázově generován na SpecvP, která je výše než SpecvP s neživotným subjektem. Celkově musí být dodržen slovosled životný subjekt následovaný neživotným objektem, aby věta s LBE byla přijatelná. Pokud je životnost obou konstituentů vyrovnaná, musí být dodržen slovosled subjekt – objekt.

Klíčová slova: extrakce levé větve, slovosled, česká syntax, fáze, linearizace, sémantika

Table of contents

List of abbreviations

ACC	Accusative (case)*
AUX	Auxiliary*
C	Complementizer
CP	Complementizer phrase
DP	Determiner phrase
F	Female (gender)*
INSTR	Instrumental (case)*
LBC	Left branch condition
LBE	Left branch extraction
LOC	Locative (case)*
M	Masculine (gender)*
N	Noun
N	Neuter (gender)*
NEG	Negation*
NOM	Nominative (case)*
NP	Noun phrase
P	Preposition
PL	Plural*
PP	Prepositional phrase
REFL	Reflective*
Spec	Specifier
SVO	Subject Verb Object
TP	Temporal phrase
UTAH	Uniformity of theta-assignment hypothesis
V	Verb
VP	Verb phrase

*Abbreviations marked with * were used for glossing of non-English material.*

1 Introduction

Czech has a so-called free word order. Yet, there are structures which seem to be stricter about their word order than others. In this thesis, we took a closer look at a phenomenon called left branch extraction. Although in English considered as a forbidden movement out of a phrase, in Czech very much possible. But under what circumstances? And what does it say about the Czech word order in general?

In the second chapter, a theoretical background for this phenomenon will be given. Firstly, left branch condition / extraction will be introduced. Although not possible in for example English, some languages do permit left branch extraction. Besides Czech, Slavic languages like Russian or Polish use these constructions. Due to the most research done on these two languages, examples of LBE will also be given from them. They will function as the basis for the analysis of Czech (some aspects are same for the languages). The next part will focus on an analysis of structures with LBE rooted in generative syntax. The phase theory and cyclic linearization will also be explained. Both theories work with the concept of spell-out. Although the analysis was introduced for Russian. As already mentioned, this language has similarities in the extraction of the left branch of an NP with Czech. The empirical evidence will later show us if the analysis is indeed applicable for Czech.

Next question which will be tackled is whether the word order is strictly a matter of syntax or if semantics plays its role. To be more precise, does only the syntactic function of constituents determine the word order or is the order also dependent on the semantic nature of the constituents? The semantic category most discussed in this thesis is the animacy (or rather the humanness). The section will explore the possibility of analysing construction in which the human constituent takes a prior position to the nonhuman constituent. Two hypotheses of such analysis will be introduced, one of them respects the uniformity of theta-assignment hypothesis (UTAH), the other relaxes the UTAH.

The third chapter will cover the empirical part of this thesis. An experiment was conducted to test our hypothesis derived from the theory. Altogether, the experiment consisted of 1 main experiment and 5 filler experiments, each examining a different aspect of constituents figuring in LBE. Besides already mentioned variables like humanness, the effect on the clause of overtness / covertness of the constituents was tested. One filler experiment also works with two types of adverbs which should behave differently in the structures, the analysis of such structures is included in this section. The participants of the experiment rated various clauses

with LBE (and filler experiment 5 without LBE) on a scale from natural to unnatural. All items were created specifically for this experiment.

The results and their impact on the theory are then discussed in the last part of this thesis.

2 Left branch extraction

2.1. Introduction to the phenomenon

What is left branch extraction (LBE) and how does it connect to Czech? To discuss this, we first need to mention the proposal of left branch condition (LBC) by Ross (1967):

"No NP that is the leftmost constituent of a larger NP can be reordered out of this NP by a transformational rule."

LBC forbids a movement of a left constituent from an NP. The condition is based on English, in which it is relevant. If a question concerning the leftmost constituent in (1) *a* is to be asked, the wh-element needs to occupy the first position of the clause. As can be seen in *b* and *c,* moving only a part of the NP to the left of its head makes the whole clause agrammatical. The question is possible only when the whole NP is fronted.

(1) (Corver, 1990)

a John saw [[[the minister's] wife's] dog]$_{NP}$.

b *Whose$_i$ did you see [[[t]$_i$ wife's] dog]?

c *Whose wife's$_i$ did you see [[t]$_i$ dog]?

d Whose wife's dog$_i$ did you see [t]$_i$?

This condition, however, does not seem to apply to Czech, where a leftmost constituent of an NP can indeed be moved. Examples in (2) show the difference between questions without LBE and with LBE in Czech.

(2)

a$_1$	[Která	dívka]$_{NP}$	čte?				*without LBE*
	which.NOM	girl.NOM	reads				
a$_2$	Která$_i$	čte	[t$_i$ dívka]$_{NP}$?				*with LBE*
	which.NOM	reads	girl.NOM				
	'Which girl reads?'						
b$_1$	[Kterou	dívku]$_{NP}$	jsi	viděl?			*without LBE*
	which.ACC	girl.ACC	be.AUX	saw			
b$_2$	Kterou$_i$	jsi	viděl	[t$_i$ dívku]$_{NP}$?			*with LBE*
	which.ACC	be.AUX	saw	girl.ACC			
	'Which girl did you see?'						
c$_1$	[Jakou	knihu]	jsi	přečetl?			*without LBE*
	which.ACC	book.ACC	be.AUX	read			
c$_2$	Jakou$_i$	jsi	přečetl	[t$_i$ knihu]?			*with LBE*
	which.ACC	be.AUX	read	book.ACC			

‘Which book did you read?’

“Která dívka”, “kterou dívku” and “jakou knihu” form NPs (as indicated in (2)). The Wh-constituent (“který” or “jaký” in the right gender and case in our examples) stands in the first position (before a clitic). Depending on the situation, “který”/”jaký” can either be moved to its place by itself (a_2, b_2 and c_2 with LBE) or as a whole NP (a_1, b_1 an c_1 without LBE). The latter mentioned is the only possibility in English, as is represented in (1) *d*. Both alternations of *a*, *b* and *c* share the same meaning.

2.2 Languages and LBE

Bošković (2008) argues that LBE is possible in languages that do not have articles, i.e., they have NP and not DP. He postulates that DP is a phase (and causes problems with the movement), NP however is not a phase. In English that has articles is thus the movement not possible (as seen in examples in (1)), other language prohibiting LBE is for example Dutch (Corver, 1990), German (examples for both in (3)) or French.

(3) *Examples from languages permitting LBE*

German (Fanselow & Lenertová, 2011)

a	[Obamas	Buch]$_i$	hab’	ich	t_i gekauft.
	Obama’s	book	have	I	bought
b	*Obamas$_i$	hab’	ich	[t_i Buch]	gekauft.
	Obama’s	have	I	book	bought

‘I have bought Obama’s book.’

Dutch (Corver, 1990)

c	*Hoe	mooi	bezat	Jan	[een	t_i schilderij	van	Dali]?
	how	beautiful	owned	John	a	picture	by	Dali

‘How beautiful picture by Dali did John own?’

d	*Hoe	interessante$_i$	heeft	Jan	[een	t_i	lezing	gegeven]?
	how	interesting	has	John	a		lecture	given

‘How interesting lecture has John given?’

An example of a construction that appears like a violation of LBC can be found in the latter language. In (4) *a* there is not the violation of LBC because “de livres” is a postverbal de-N’ phrase. For it to be a question with LBE, the structure would have to look like *b*, which is indeed not grammatical in French.

(4) Chaves (2021)

a Combien$_i$ a-t-il vendu [t_i de livres]?[1]
how many has-he sold of books
'How many books did he sell?'

b *Quels$_i$ avez-vous acheté [t_i livres]?
how many have-you bought books
'How many books have you bought?'

Czech, however, being an articleless language meets the requirement for LBE. Other Slavic languages permitting LBE are for example Polish (Corver, 1990), Serbo-Croatian (Bošković, 2008) and Russian (Bošković, 2008; Bondarenko & Davis, 2021). Chaves (2021) further argues that without articles (i.e., determiners) the phrase from which the left branch is extracted could be independent similarly to the French "de livres", but he does not give any evidence for this claim. Examples of LBE from mentioned languages are in (5).

(5) a Jaką$_i$ kupiłeś [t_i książkę]? *Polish (Chaves, 2021)*
which bought book
'Which book did you buy?'

b Ètogo$_i$ devočka pogladila [t_i kota]. *Russian (Bondarenko & Davis, 2021)*
this.ACC girl stroked cat.ACC
'The girl stroked this cat.'

c Skupa$_i$ je vidio [t_i kola]. *Serbo-Croatian (Bošković, 2008)*
expensive is seen car
'He saw an expensive car.'

2.3 LBE in Czech

If looked upon from a different view, constructions with LBE are nonprojective constructions created by the division of an NP. In a paper discussing the Czech nonprojective constructions in the Prague Dependency Treebank Hajičová, Havelka, Sgall, Veselá and Zeman (2004) state that the percentage of clauses with divided nominal groups is about 11 % and fronted wh-elements about 1.6 % (however it is not clear whether all the findings are strictly cases with LBE). As is apparent from these percentages, constructions with LBE are quite uncommon (but possible).

Left branch cannot be extracted if the NP is a complement of a preposition. Question in (6) *c* is not grammatical. To move the left branch of the NP "kterém domě" to a higher position (i.e., to a fronter position), the head of the PP has to move to the front as well (example in *b*). The whole

[1] Corver (1990) actually uses this sentence as an example of left branch reordering from NP. He describes the phenomenon is possible in some non-Slavic languages only to some extent.

PP can also be fronted to form the question (in *a*). (5) *d* shows another ungrammatical variation of the clause, the NP cannot be moved while its sister (the head of PP) is left in its base position.

(6) a [V kterém domě]$_{PP}$ bydlel?
in which.LOC house.LOC lived

b [V kterém]$_i$ bydlel [t_i domě]$_{PP}$?
in which.LOC lived house.

c *Kterém$_i$ bydlel [v [t_i domě]$_{NP}$]$_{PP}$?
which.LOC lived in house.LOC

d *[Kterém domě]$_i$ bydlel [v t_i]$_{PP}$?
which.LOC house.LOC lived in

'In which house did (he) live?"

Corver (1990) illustrates these same conditions also for Polish. (7) *a* corresponds with (6) *d*, (7) *b* then with (6) *c*, which shows the impossibility of extracting the left branch of NP from a PP without the fronting of P. For (7) b, the problem is again the extraction of the left branch "jakim" from the NP that is complement of a PP.

(7) *adapted*

a *Kim$_i$ rozmawiałes [z t_i]?
who.INSTR talked with
'Who did you talk with?'

b *Jakim$_i$ on mieszka [na [t_i piętrze]$_{NP}$]$_{PP}$?
which.LOC he lives on floor.LOC
'On what floor does he live?'

Examples of a more complex left branch of an NP complementing a PP are mentioned by Veselovská (2021):

(8) *adapted*

a [Za čí]$_i$ jste bojovali [t_i nezadatelná práva]?
for whose.ACC be.AUX fought indisputable.ACC rights.ACC

b [Za čí nezadatelná]$_i$ jste bojovali [t_i práva]?
for which.ACC indisputable.ACC be.AUX fought rights.ACC
'For whose indisputable rights did you fight?'

c [Za jaká ta jeho]$_i$ jste bojovali [t_i práva]?
for which.ACC those.ACC his.ACC be.AUX fought rights.ACC

d *[Za jaká]$_i$ jste bojovali [ta jeho t_i práva]?
for which.ACC be.AUX fought those.ACC his.ACC rights.ACC
'For which (the) rights of his did you fight?'

The ungrammaticality of (8) *d* is caused due to the possessive "ta jeho" being left on the base position. As is seen in *a* and *b*, where the possessive is moved to the front of the clause, the

AdjP follows the possessive, thus "za čí" can be fronted by itself in *a*.[2] To ask about the nature of the rights, it is needed to move the possessive as well, because "jaká" by itself is not the initial part of NP.

All of the above examples from Czech are questions. Left branch of an NP in Czech, however, is not extracted solely for the creation of questions (although this purpose is probably the most frequent one). In an indicative clause, an adjective can be fronted from an NP as an indicator of a stress (marked in majuscules in (9)).

(9) a [Altový saxofon] jsem si vždy přála.
alto.ACC saxophone.ACC be.AUX REFL always wished
'I always wished for an alto saxophone.'

b ALTOVÝ$_i$ jsem si vždy přála [t_i saxofon].
alto.ACC be.AUX REFL always wished saxophone.ACC
'I always wished for an ALTO saxophone.'

Fanselow & Lenertová (2011) give examples of indicative clauses with LBE given as answers to a question (which is actually also an example of a structure with LBE from an object).

(10) *adapted*
Q Jakou$_i$ jí koupil [t_i růži]?
what.ACC her.DAT bought rose.ACC
'What kind of a rose did he buy her?'

a [ČERvenou růži]$_i$ jí koupil t_i.
red.ACC rose.ACC her.DAT bought.

b ČERvenou$_i$ jí koupil [t_i růži].
red.ACC her.DAT bought rose.ACC

c Koupil jí [ČERvenou růži].
bought her.DAT red.ACC rose.ACC
'He bought her a red rose.'

All (10) *a*, *b* and *c* are valid answers for the given question. In the first clause, the whole phrase is moved to the front, example in *c* leaves the phrase on the base position (no movement happens). The second one has the left branch extracted from the NP that is an object. Interestingly, if the question and the context changes to (11), the plausible reactions can be found in (10) *a* and *c*.

[2] The indicative clause with an unmarked word order in (I) also shows the possessive as the leftmost constituent of the NP phrase dominated by the PP.

(I) Bojovali jste za ta jeho nezadatelná práva.
fought be.AUX for those.ACC his.ACC indisputable.ACC rights.ACC
'You fought for the indisputable rights of his.'

(11) *Fanselow & Lenertová (2011)*

Karel invited Hana to the cinema and Jan brought he a white rose.

What did Emil buy?

Last category of structures discussed in this chapter will be the clauses where the verb is transitive and both subject and object are present in the clause (in other words, both constituents are overt). These create a majority of items in our experiment further discussed in chapter 3.

(12) *b* and *c* illustrate the two possible left branch extractions derived from the clause in *a*.

(12) a Malý chlapec si koupil modré pero.
little.NOM boy.NOM REFL bought blue.ACC pen.ACC
'A little boy bought a blue pen.'

b *LBE from object*
Jaké$_i$ si malý chlapec koupil [t_i pero]?
which.ACC REFL little.NOM boy.NOM bought pen.ACC
'Which pen did a/the little boy buy?'

c *LBE from subject*
?Jaký$_i$ si modré pero koupil [t_i chlapec]?
which.NOM REFL blue.ACC pen.ACC bought boy.NOM
'Which boy did buy a/the blue pen?

If looked at all the previous examples plus the further three questions presented in (13), the examples with LBE from an object seem to dominate (even in the other languages beside Czech).

(13) a Jakou$_i$ by Jan dal [t_i knížku] *adapted from*
which.ACC.F be.AUX Jan.NOM give book.ACC.F *Corver (1990)*
Markovi?
Marc.DAT
'Which book would Jan give to Marc?'

b Jakou$_i$ čte Petr [t_i knihu]?
which.ACC.F reads Petr.NOM book.ACC.F
'Which book does Peter read?'

c Kolik$_i$ četl [t_i knih od Nesbøho]? *adapted from*
How many read books by Nesbø *Veselovská (2021)*
'How many books by Nesbø did he read?'

LBE from a subject is exemplified only in (2) *a* which we repeat in (14) for better orientation (the case of (12) *c* will be discussed shortly). In this case, the verb "číst" 'to read' does not need an object. When questioning the nature of the subject "dívka", structure with LBE can be used without any restrictions, nothing is intervening the extraction, the clause is grammatical.

(14) a_1 [Která dívka]$_{NP}$ čte? *without LBE*
which.NOM girl.NOM reads

a_2 Která$_1$ čte [t_1 dívka]$_{NP}$? *with LBE*
which.NOM reads girl.NOM

'Which girl reads?'

What about the question in (12) *c*? As native speakers of Czech, we perceive this clause as far worse than (12) *b*, if not even ungrammatical. The only difference between (12) *c* and (14) a_2 is the presence of the intervening constituent (more precisely the presence of an intervening object). Bondarenko & Davis (2021) stumbled across the influence of scrambling on the possibility of LBE in Russian:

(15) *adapted*

a Kota$_i$ [èta devočka] pogladila t_i.
cat.ACC this.NOM girl.NOM stroked

b *Èta$_i$ kota$_k$ [t_i devočka] pogladila t_k.
this.NOM cat.ACC girl.NOM stroked
'This girl stroked the cat.'

c Ètogo$_i$ devočka pogladila [t_i kota].
this.ACC girl stroked cat.ACC
'The girl stroked this cat.'

(15) *a* is a clause without LBE, the only movement that happens is the move of NP "kota" to the front, the phrase functions as an object. Subject stays on its base position. But if the same movements is to happen in a clause with extraction, the whole clause is ungrammatical. In (5) *b* a similar construction was exemplified, in (15) added as *c.* Here the extraction happens from an object. The intervening constituent is a subject. Yet, the clause is perfectly fine and plausible. This is also true for the Czech material. The problem therefore seems not to be caused by the intervention of a constituent in the extraction.

So, what is the difference between the clauses that makes some of the LBE examples ungrammatical? What is the distinction between the presented extractions from an object and those from the subject?[3] It is the presence of scrambling. It seems that the extraction is not possible with the scrambled word order. As in our example in Czech ((12) *c*), the extraction from the subject is intervened by the object which had to move on this position from its base lower in the structure. Because a subject is already higher in its base position than an object in

[3] This distinction will be further complicated in 2.5, at this stage we will work with the basic word order being subject – object.

a structure, the extraction from an object with an intervening subject is possible due to the lack of scrambling.

With this in mind, the question now is how to describe all of the observations in theory.

2.4 Analysis

For our analysis, we will work with the phase theory and cyclic linearization.

2.4.1 Phase theory

To describe the structure of clauses, Chomsky (2000) suggests the phase theory. The clause is divided into phases. If some part of a phase is needed in another phase, the part is only accessible from the edge of the base phase (see 16).

(16) Phase-Impenetrability Condition

In phase a with head H, the domain of H is not accessible to operations outside a, only H and its edge are accessible to such operations.

In other words, to be able to work with a constituent from phase "x" in a higher phase "y" the constituent first needs to be moved to the edge of phase "x", from this edge the constituent can be then moved in phase "y". If a constituent stays in a phase, it undergoes the operation spell-out, by which the phase, more precisely the complement of the phase head, is sent to logical and phonological form "and thus by hypothesis, out of the syntactic derivation" (Bondarenko & Davis, 2021). After this, the constituent is no longer able to move to other phases. This is called successive-cyclic movement.

Figure 1 shows a non-successive-cyclic phase exit (*a*) and a successive-cyclic phase exit (*b*). In this case, vP is a phase. Constituents which are no longer available for movement beyond vP are marked with the blue line, anything lower than it is prohibited from beyond-phase movement (in figure 1, VP and its daughters). The edge of the phase can be found on v and DP_1. Thus, in case of a non-successive-cyclic phase exit in *a*, V cannot be moved from its base place to T because V is not on the edge of vP (it is below the blue line). In the case of a successive-cyclic phase exit (*b*), V is first moved to v (which is possible), from v it can then be moved further to T due to its position at the edge of vP.

a b

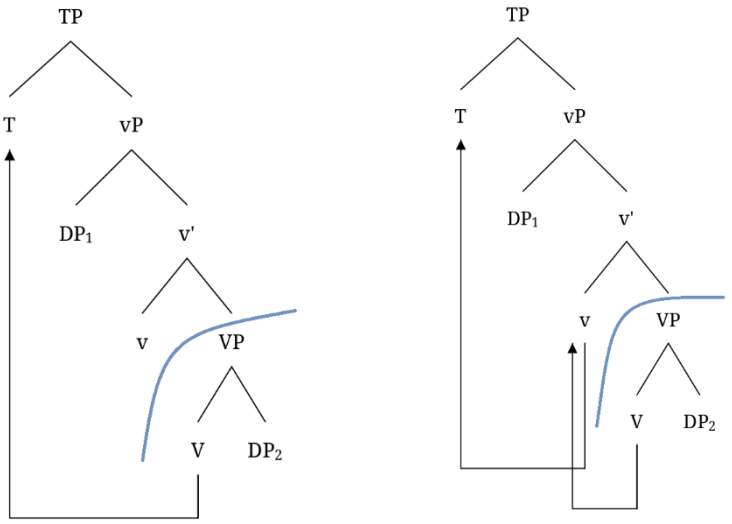


Figure 1 – visualisation of cyclic movement, adapted and altered from Biskup (2017b)

2.4.2 Cyclic linearization

Cyclic linearization (Fox & Pesetsky, 2005) also deals with the operation of spell-out. Linearization which handles the word order at spell-out happens phase by phase. After spell-out follows movement of constituents. For a movement to be successive-cyclical, the word order has to be the same after all spell-outs, otherwise, the constituent would be present twice in one clause, more precisely the linearization would cause a problem in the phonological form.

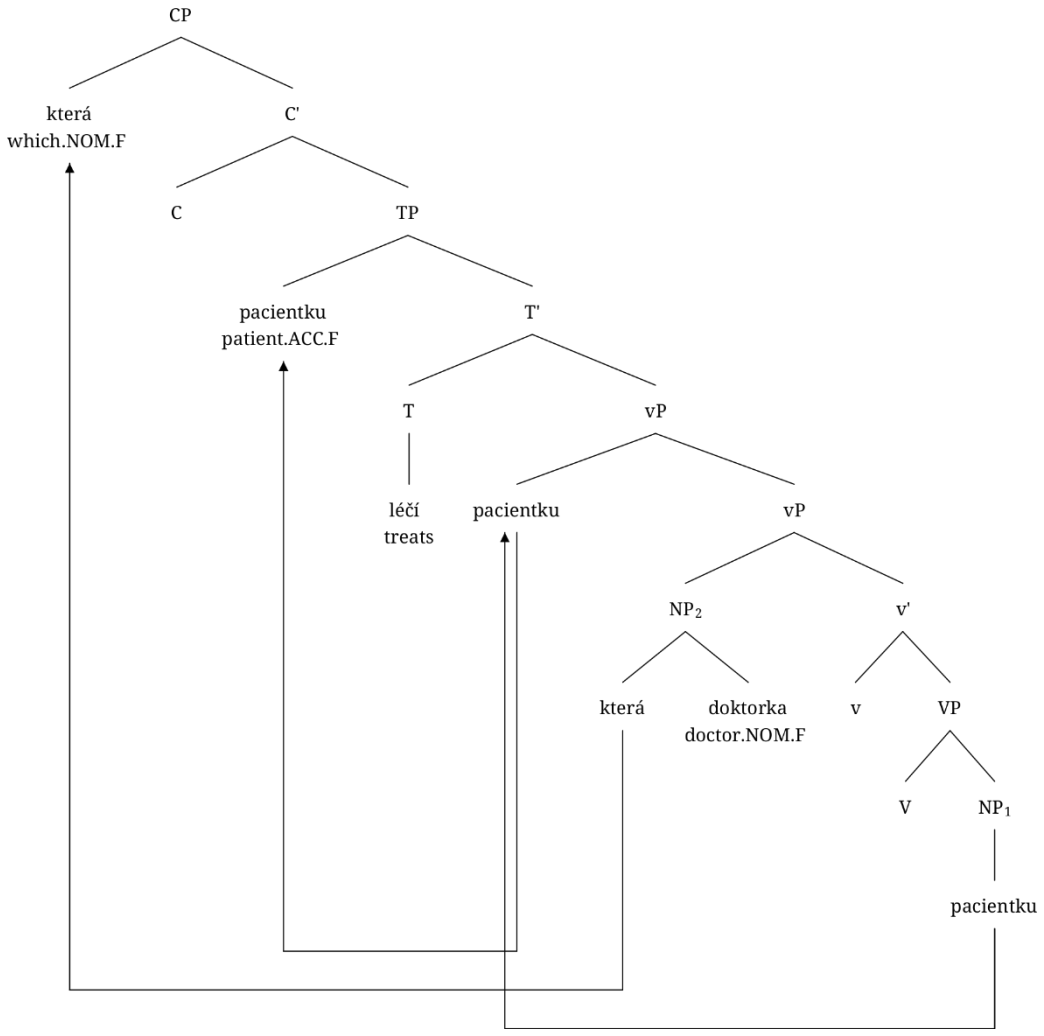
Bondarenko & Davis (2021) point out that the contradiction between linearization before and after spell-out can be solved without successive-cyclical movement. If constituent "y" is to be moved from a place that is not an edge of a phase to a higher phase, it is needed to also move the constituent "x" which is between the higher phase and the base position of constituent "y" (see (17)). By doing so, the word order will be the same after the spell-out of vP and after the linearization of CP.

(17) *adapted from Bondarenko & Davis (2021)*

a [x y [t_x t_y]$_{ZP[phase]}$]$_{FP[phase]}$

b *[y [x t_y]$_{ZP[phase]}$]$_{FP[phase]}$

To demonstrate this, we use examples (figures 2 and 3) from our experiment which is further discussed in chapter 3.

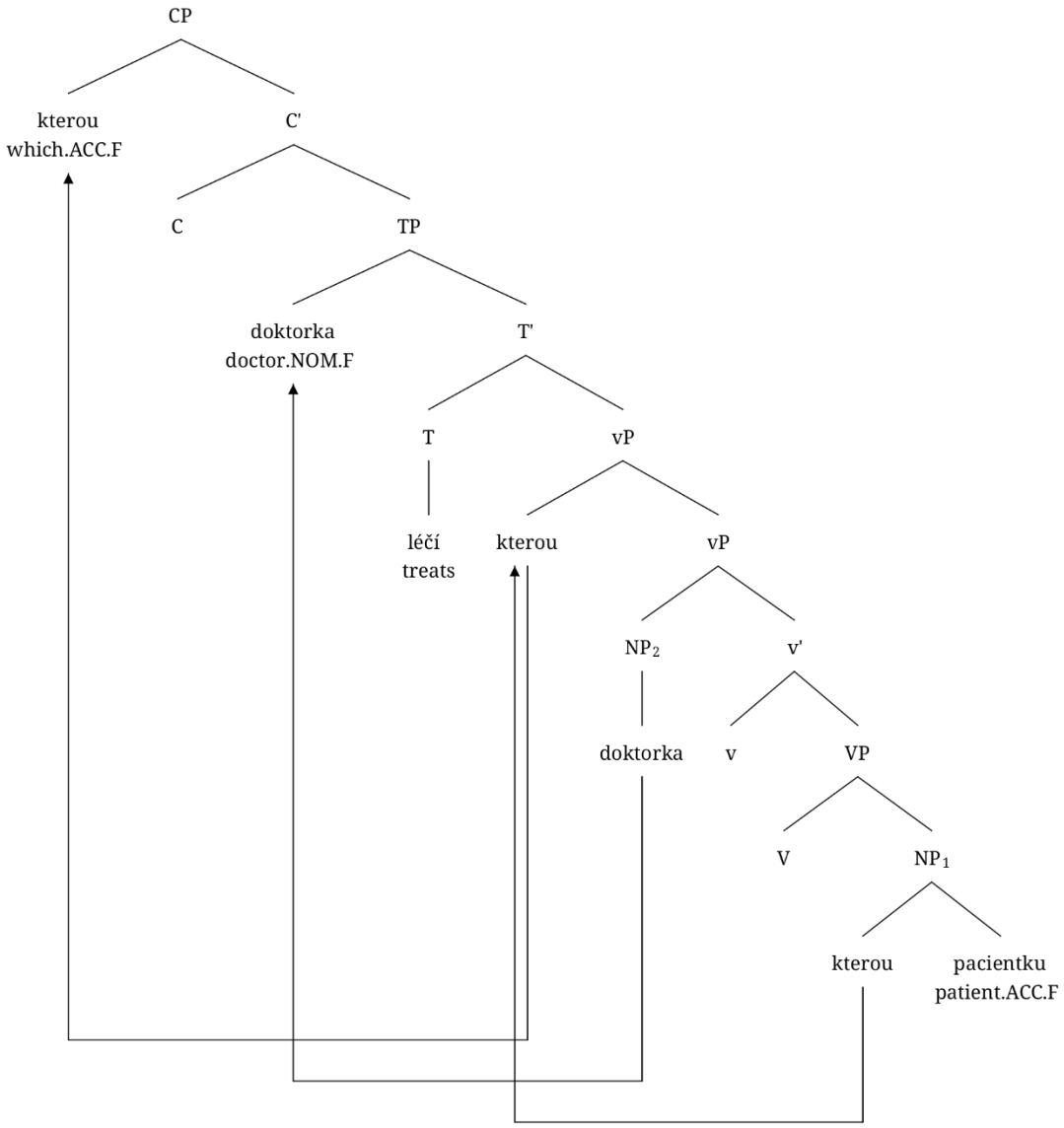


Která pacientku léčí doktorka?
'Which doctor treats a/the patient?'

Figure 2 – example of a linearization problem

NP_1 "pacientku" is firstly moved from a non-edge base position into SpecvP, then to SpecTP. In NP_2 left branch is extracted and moved to SpecCP. The head of NP_2 remains in its base position. The word order after linearization of vP is "pacientku" – "která" – "doktorka", after the linearization of CP "která" – "pacientku" – "doktorka", and thus a linearization problem occurs. The clause "Která pacientku léčí doktorka?" would be acceptable by this theory if the whole NP_2 (since it is in between the SpecvP and the base position of NP_1 moved without successive-cyclical pass of the edge) would move to SpecCP. In this case, the clause would be "Která doktorka léčí pacientku?" (and without LBE). This is the only possible way for the clause to be grammatical. At first sight, it seems that the linearization conflict would not be present, if the left branch "která" first moved on a higher position to the one containing "pacientku" and remaining still in SpecvP. In other words, it would mean that the constituent moves from SpecvP to SpecvP, which is forbidden (Ko, 2007) .The problem with this movement

is that SpecvP is already outside of the search domain (c-command domain) of little v, thus, there is nothing that could trigger the Spec-to-Spec movement.



Kterou doktorka léčí pacientku?
'Which patient is treated by a/the doctor?'

Figure 3 – example of a coherent linearization

This clause, on the other hand, shows consistency in linearization. LBE happens in NP_1, the pronoun "kterou" moves to SpecvP, from there to SpecCP, therefore avoiding the movement through the edge of the phase. NP_2 is moved from its base position to SpecTP. The word order after the linearization of vP is "kterou" – "doktorka" – "pacientku", which is the same as the word order after the linearization of CP (i.e., same as the word order of the final clause).

The problem generally seems to be the scrambling of the constituents in combination with the extraction (as was mentioned in 2.3). If no scrambling happens (the word order stays the same), the extraction (does not matter whether LBE from an object or a subject) is possible. If there is

scrambling which would interference with the word order (the word order would change throughout the process), the extraction (again no matter the type) is not possible.

2.5 Adding humanness to consideration

Czech is categorized as an SVO language (Sgall et al., 1980; Uhlířová, Kučerová, 2017 and others)), the canonical, unmarked word order is an object following a subject. On the other hand, Czech has quite a flexible word order (more precisely, free word order). As was exemplified in (9), one reason for a noncanonical word order can be a stress of a certain constituent.

If we come back to the basic word order, Titov (2012) argues that the semantic nature of constituents has precedence in constructing the word order, rather than their syntactic function. On an example from Russian (2012), Titov shows the precedence of an animate constituent to an inanimate one, even though the animate one is indirect object and the inanimate a direct object which should have (syntactically) prior position to the other one. After asking "What did Ivan do?", a more natural response is found in *a* than in *b*. In the latter clause, the word order is an inanimate object followed by an animate direct object. This points to the direction, that the semantic nature of the constituent is the first variable in decision making of what the neutral word order looks like.

(18) *Titov (2012)*
Q Čto Ivan sdelal?
'What did Ivan do?'
a Ivan peredal agentu pis'mo.
Ivan handed agent.DAT letter.ACC
'Ivan handed the/a letter to the/a agent.'
b Ivan peredal pis'mo agentu.
Ivan handed letter.ACC agent.DAT
'Ivan handed the letter to an agent.'

If we look at Czech, the tendency seems to be the same. After being asked the question "Co se stalo?" ('What happened?') the ordering in (19) *a* is a better option for response than the clause in *b*. The first clause has object – subject worder order, but the animate constituent is on the first place, inanimate on the second, whereas *b* is ordered nonhuman subject – human object. The order of words in *a* is therefore a neutral order.

(19) Q Co se stalo?
'What happened?'
a Běžkyni srazilo auto. *OVS, human – nonhuman*
runner.ACC.F knocked down.N car.NOM.N

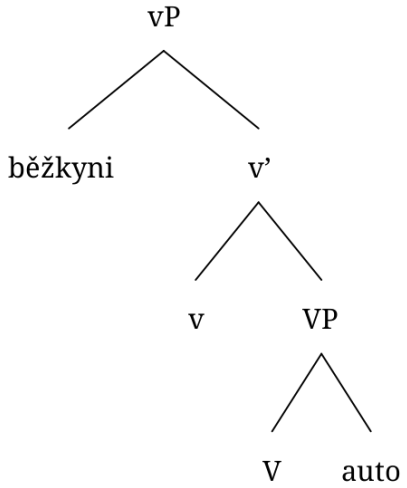
b Auto srazilo běžkyni. *SVO, nonhuman – human*
car.NOM.N knocked down.N runner.ACC.F

'A/the car knocked down a/the runner.'

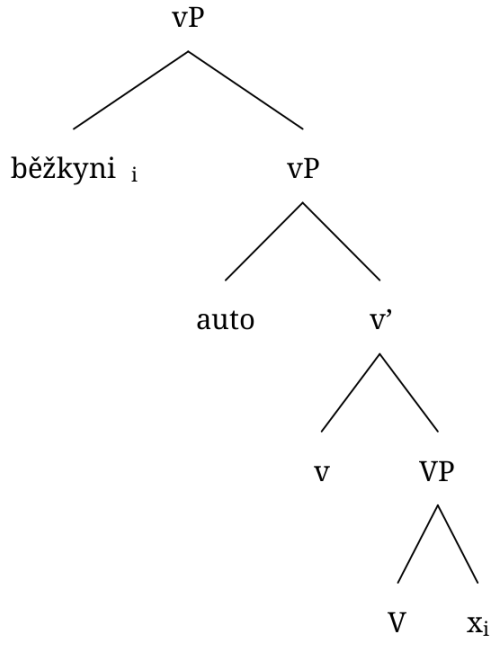
The humanness of the constituent is dominating the syntactic function, thus an object that is semantically human (or rather animate) precedes the subject that is nonhuman (inanimate). If the animacy of constituents in a clause is the same, then is the word order based on their syntactic function.

But how is this factor possible if the subject is based on a higher position than the object as can be seen in Figures 2 and 3? To answer this question, the base-generated approach to scrambling is needed to be introduced. Following previous work (e.g., Fanselow 1993, 2001, 2003; Neeleman 1994), Titov (2012) proposes that in cases like example in (19) where the subject denotes a non-human and the object a human, the object is base-generated in front of the subject. As for how this can be achieved, the structures in (20) provide a visual representation for better understanding (taken from (19) *a,* setting aside the position of the verb).

(20) a



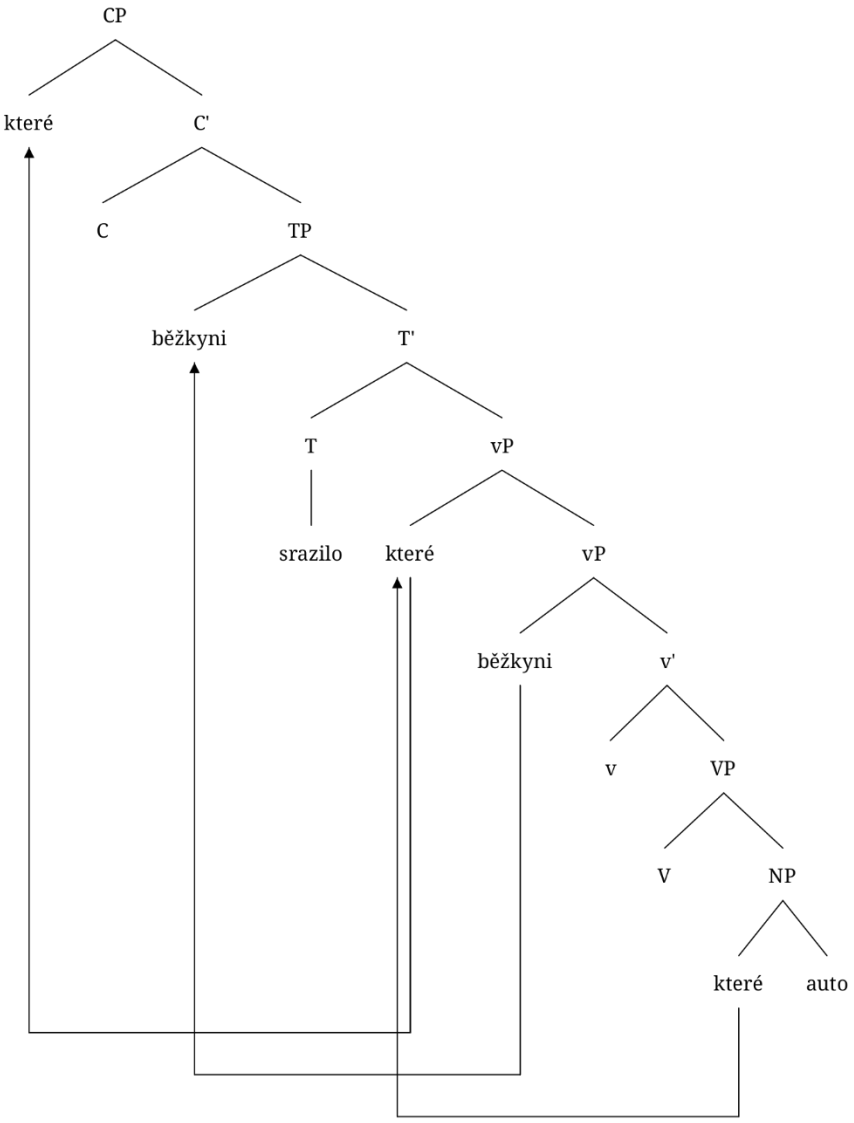
b



The first way of achieving the base position of a human object to be higher than the base of nonhuman subject is by relaxing the uniformity of theta-assignment hypothesis (UTAH; Baker 1988). The subject can then be let to be base-generated in the complement of V (which is the canonical object position) and the object in SpecvP (the canonical subject position). This option is employed in e.g., Fanselow (2001).

The second way, on the other hand, respects the UTAH. The subject is therefore always base-generated in SpecvP. The object itself has a base position on a higher SpecvP (thus is on a fronter position than the subject), from there the object would bind a semantic variable (i.e., not a trace) in the complement of V. This was hinted at in Fanselow (2003).

What predictions do these two theoretical and analytical options make about the possibility (availability) of LBE in the configurations where the humanness of the constituents is non-canonical? They give rather contradictory predictions. Figure 4 illustrates the first option (the "swapping" of base positions of subject and object).

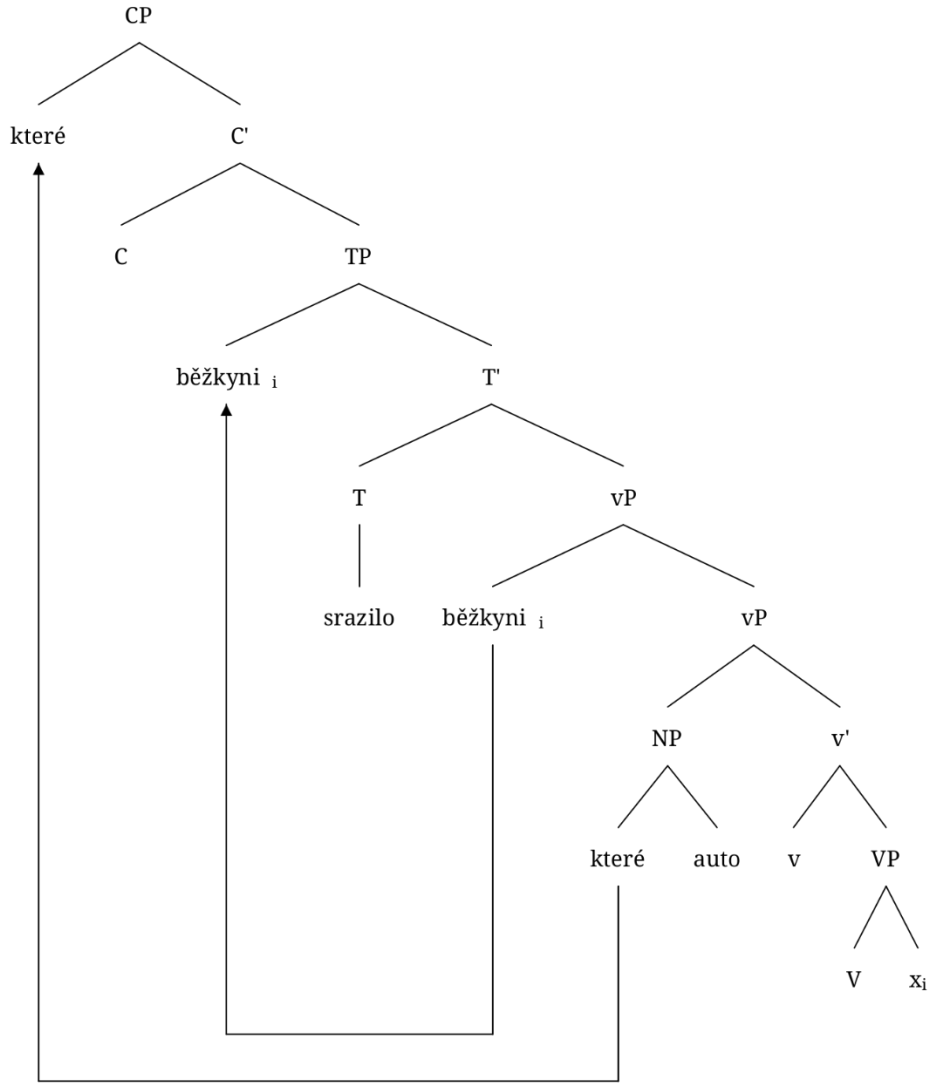


Které běžkyni srazilo auto?
'Which car knocked down the runner?'

Figure 4 – LBE from the subject in a question created from (19) a with the representation of base positions of a human object and a nonhuman subject found in Fanselow (2001)

This structure is similar to the one found in figure 3, or in other words, as the extraction from an object over a subject in a canonical word order base on syntactic function. “Které” ‘which’ is moved from its base position to SpecvP. After the linearization of vP, the word order is “které – běžkyni – auto”. ‘Which’ is then moved to SpecCP, the object “běžkyni” ‘runner’ is moved to SpecTP. After the linearization of CP, the word order stays the same as was after the prior spell-out, i.e., “které – běžkyni – auto”. This means that no linearization problem occurs and that the structure should be possible, as was the case with the clause in figure 3.

The second option which respects the UTAH can be found illustrated in figure 5.



Které běžkyni srazilo auto?
‘Which car knocked down the runner?’

Figure 5 – LBE from the subject in a question created from (19) a with the representation of base positions of a human object and a nonhuman subject found in Fanselow (2003)

In this case, the NP which is functionally the subject is already in vP (including the left branch “které” ‘which’). The word order after the linearization of vP looks as following: “běžkyni –

které – auto". "Které" moves to the front (to SpecCP) and "běžkyni" to SpecTP. The word order after second linearization is "které – běžkyni – auto". The linearization at the CP level is then in conflict with the first linearization of vP. As in the case of extraction from human subject over an object (see section 2.4.2), it is not possible to avoid the linearization conflict by moving 'which' from within the nonhuman subject in the SpecvP to a higher SpecvP (Ko, 2007).This explanation of the representation of a word order based on humanness therefore forbids the possibility of LBE from the nonhuman subject.

All in all, the first hypothesis (as shown in the fourth figure) suggests that the extraction of the left branch should not be possible unless the word order stays canonical. In other words, given the word order a nonhuman subject – a human object the whole clause with LBE is ought to be unacceptable. The second hypothesis (figure 5), on the other hand, states that only the extraction from a nonhuman object over a human subject is possible.

The proposed basic word order in Czech is therefore a human subject – a nonhuman object, if the humanness is balanced, the word order is subject – object.

3 Experiment

To test the claims derived from the theoretical part, we created an experiment which will be discussed in this chapter.

3.1 Design

The experiment consists of 1 main experiment and 5 filler experiments. All items are in the form of an indirect question, thus include a main clause followed by a subordinate clause. Our focus is on the latter one which contains LBE (except for F5, see 3.1.6) and is introduced by the pronoun "který" ('which') in the right gender and case. LBE is in all cases either from an object over a subject or from a subject over an object. The main clause was added for context and to make some of the clauses more plausible, each item has one main clause for all conditions (exceptions discussed in 3.1.2).

We tried to use different noun phrases and verbs in all the items to reduce priming, though some of the matrix verbs appear more than once due to the fact that only a section of verbs can be used as a matrix verb in indirect questions. On the other hand, this should not be a problem since these verbs are generally quite common (e.g., say, ask, know), especially in the context of an indirect question, and should not interfere with the acceptability of the clauses in general. The nouns figuring in LBE (either as the noun in NP from which is extracted or as the intervening constituent) are mostly female, alternatively other gender with different forms of NOM and ACC, because it is clearer for the recipient to see which parts belong together (and avoiding a garden path).

The verbs of the LBE part of the clause were selected based on their valency. Firstly, they needed to be transitive verbs to test both types of extractions. Secondly, both subject and object have to be able to be semantical human and nonhuman (some verbs were discarded due to the fact that their nonhuman constituent could only be an animal and not a thing, e.g., "lekat" 'to startle').[4] The last factor was the interchangeability of an object NP and a subject NP, precisely the ability of two nouns to be both an object and a subject with the other one being the other constituent.

[4] In our case, nonhuman constituent is always a thing, animals were excluded due to their animacy status. The difference between an animal constituent and a human or a thing on the word order is open for future research.

3.1.1 Main experiment

The design of the main experiment is 2 x 2, variables being the constituent order (i.e. the type of extraction) and the order of humanness of nouns, in this case either human – nonhuman or nonhuman – human. It consists of 24 items. The nouns of the subordinate clauses were chosen so that they can function as a subject as well as an object with the same verb, thus all four conditions in one item use the same words.

	subject – object	*object – subject*
human – nonhuman	c	b
nonhuman – human	a	d

Table 1 – Main experiment – 2 x 2 design

3.1.2 Filler experiment 1 – pronouns

In the first filler experiment, our goal was to look at how the overtness of the intervening constituent interacts with the naturalness of the whole clause. The experiment includes 8 items with the design 2 x 2 x 2. The variables are the same as in the main experiment (see 3.1.1) plus the category of the intervening constituent. The constituent is either a noun (more precisely NP) or a pronoun (in the case of a covert subject, the pronoun is omitted as it is common to do so in Czech). The subject and object of one item are again interchangeable. After creating the first version of items, we decided to mention the intervening constituent (in its overt form) in the main clause, so that the covert constituent would more likely be interpreted as a human or a nonhuman depending on the noun. Without this, we found the covert form to be naturally interpreted as a human which we do not want. This leads to one item having two versions of the main clause to capture both NPs (each clause for four conditions).

	nonhuman – human	*human – nonhuman*
subject – object	a	e
object – subject	g	c
pro – object	b	f
pro - subject	h	d

Table 2 – Filler experiment 1 – 2 x 2 x 2 design

3.1.3 Filler experiment 2 – humanness balanced

The second filler experiment consists of 8 2 x 2 items. The variables are the constituent order and the humanness of both constituents so that the NPs have balanced humanness. Because of this, two conditions of one item share the same human subject and human object, the other two conditions have the same nonhuman subject and nonhuman object. The main clause remains identical for all four conditions.

	subject – object	*object – subject*
nonhuman – nonhuman	a	b

human – human	c	d

Table 3 – Filler experiment 2 – 2 x 2 design

3.1.4 Filler experiment 3 – definiteness of the intervening constituent

With the design of 2 x 2, the third filler experiment focuses on the nature of the intervening NP. The first variable is once more the order of constituents (all the NPs are human). For the second variable, the category of definiteness of the intervening constituent was chosen. We came up with contexts where in a group of people is one person unique (e.g., the context of an orchestra, where there are many players but only one conductor). The unique person should be semantically more definite (since we can be quite sure about whom we hear when the person is just one) than someone from the group (using our previous example, a conductor is more definite than a violin player in an orchestra with twenty other violin players).

	subject – object	*object – subject*
definite intervening const.	a	b
indefinite intervening const.	c	d

Table 4 – Filler experiment 3 – 2 x 2 design

3.1.5 Filler experiment 4 – adverbs

All of the above-mentioned experiments (including the main experiment) have a subordinate clause constituted of a subject, an object and a verb. In the fourth filler experiment, an adverb was added to the LBE clause. It consists of 8 items, four of which include the extraction from an object, the other four extraction from a subject. Each item has 2 conditions depending on the adverb used in the clause, either an adverb of manner or time. In the items, where the extraction from an object is present, is the subject omitted. Clauses with the extraction from a subject have an intransitive verb (in all the other cases only transitive verbs were used) since we do not change the type of extraction in these cases and our focus is on the type of the adverb. Each item has, again, one main clause for both conditions.

	object – subject (items 1–4)	*subject – object (items 5–8)*
adverb of manner	a	a
adverb of time	b	b

Table 5 – Filler experiment 4 – 2 x 2 design

3.1.6 Filler experiment 5 – control

The last filler experiment consists of 24 items. These clauses do not contain LBE. Half of the items were created to correspond with the standard word order, the other half are disturbed by a nonstandard word order (in most cases a wrongly situated clitic, i.e., the clitic is not on the Wackernagel's position). This filler experiment was added so that the participants would not be overloaded with clauses with LBE and to filter the participants (see 3.3.2). The items were

created to have a similar form as the items with LBE from the main experiment and the other filler experiments.

3.2 Predictions

As we are dealing with multiple hypotheses, this section will be divided into parts, each focusing on a different one. Filler experiment 4 will be discussed separately in section 3.2.5. The predictions from sections 3.2.1 to 3.2.4 are also visualised in table 6 below.

3.2.1 Word order based on syntactic function

Following the basic (canonical) word order in Czech to be solely based on the function of the constituents, and thus being subject – object (in other words, the animacy status of the constituents is irrelevant for the creation of the word order), the prediction is that the extraction from an object over a subject is always possible. The extraction from a subject over an object is not permitted. The reason for it to be ungrammatical is that the word order in the clause would be object – subject, which would have to be achieved by scrambling. Since the analysis by Bondarenko & Davis (2021) states that LBE is only possible without scrambling, this extraction would violate this presupposition.

This means that conditions *a* and *c* in the main experiment should be rated higher than conditions *b* and *d*. In the first filler experiment conditions *a*, *b*, *e,* and f are predicted to be more natural than conditions *c*, *d*, *g*, and *h*. The items in both filler experiment 2 and 3 are supposed to be perceived as better in conditions *a* and *c* than in *b* and *d*.

3.2.2 Word order based on semantics

Other possibility of looking at the potential ordering of constituents in a clause is through their semantics. In our experiment, we mainly focused on the role of the humanness of the constituents (based on the observations found in Titov (2012) and Jasinskaja & Šimík (in print)). As was discussed in chapter 2.5, there are two possible ways of analysing the clause where a human object precedes a nonhuman subject.

The first hypothesis (in the spirit of Fanselow, 2001) illustrated in (20) *a* is where the subject is base-generated in the canonical position of an object and vice versa. This analysis predicts that extraction should be possible from a nonhuman over a human, regardless of their syntactic functions. Thus, conditions *b* and *c* of the main experiment should be rated as more natural than conditions *a* and *d*. For the first filler experiment, conditions *c*, *d, e,* and *f* are supposed to be rated more acceptable than conditions *a*, *b*, *g*, and *h*.

The second hypothesis (in the spirit of Fanselow, 2003) where both nonhuman subject and human object are based in SpecvP (the latter on a higher position, see (20) *b*) predicts that extraction should be natural only in a single condition of the main experiment, namely from nonhuman objects over human subjects, i.e., condition *c*. All the other conditions involve a linearization conflict. In the first filler experiment, conditions *e* and *f* should also be rated higher than the other conditions.

In the second and third filler experiments, the humanness of the constituents is balanced. In these cases, it is the syntactic function that determines the word order. The predictions for both hypotheses are same for filler experiment 2 and 3 as mentioned in 3.2.1, i.e., conditions *a* and *c* should be more natural than conditions *b* and *d*.

Filler experiment 3 works with the definiteness of NP. This hypothesis based on the semantics of the constituents is that a definite NP should precede the less definite (nondefinite) NP. Conditions *a* and *b* based on this are predicted to be more natural than *c* and *d*.

3.2.3 Suggested basic word order

For this hypothesis, we consider the canonical word order being subject human – object nonhuman. Due to the design of the experiment, this predicts results for the main experiment and the first filler experiment.

In the main experiment, condition *c* is the only condition following the canonical word order and should therefore be the most acceptable. Conditions *a*, *b*, *d* are supposed to be unnatural. In the first filler experiment, conditions *e* and *f* are predicted to be more natural than the rest (conditions *a*, *b*, *c*, *d*, *g*, *h*).

3.2.4 Effect of overtness

In the first filler experiment, one of the variables is the overtness, or rather the covertness of the intervening constituent. The prediction is that clause with a covert form of the constituent should be rated more acceptable than ones with the overt form.

Conditions *b*, *d*, *f* and *h* are predicted to be rated higher than their overt opposites in conditions *a*, *c*, *e* and *g*.

	main experiment				*F1*								*F2*				*F3*			
	a	b	c	d	a	b	c	d	e	f	g	h	a	b	c	d	a	b	c	d
syntactic function	✓	x	✓	x	✓	✓	x	x	✓	✓	x	x	✓	x	✓	x	✓	x	✓	x

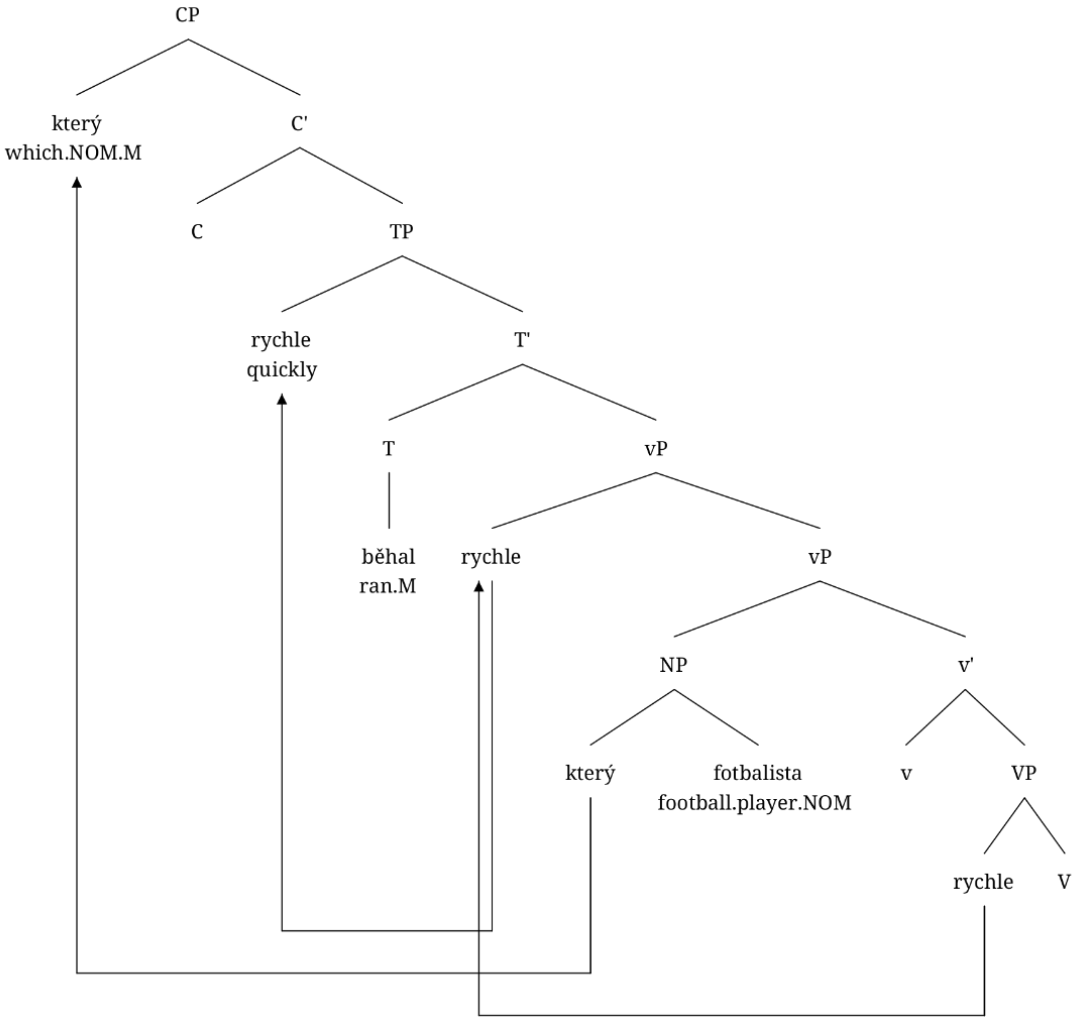
semantics in the spirit of Fanselow (2001)	x	✓	✓	x	x	x	✓	✓	✓	✓	x	x	✓	x	✓	x	✓	x	✓	x
semantics in the spirit of Fanselow (2003)	x	x	✓	x	x	x	x	x	✓	✓	x	x	✓	x	✓	x	✓	x	✓	x
definiteness of NP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	x	x
subject hum – object nonhum	x	x	✓	x	x	x	x	x	✓	✓	x	x	-	-	-	-	-	-	-	-
overtness	-	-	-	-	x	✓	x	✓	x	✓	x	✓	-	-	-	-	-	-	-	-

Table 6 – predictions for the main experiment and the first, second and third filler experiments

3.2.5 Type of adverb

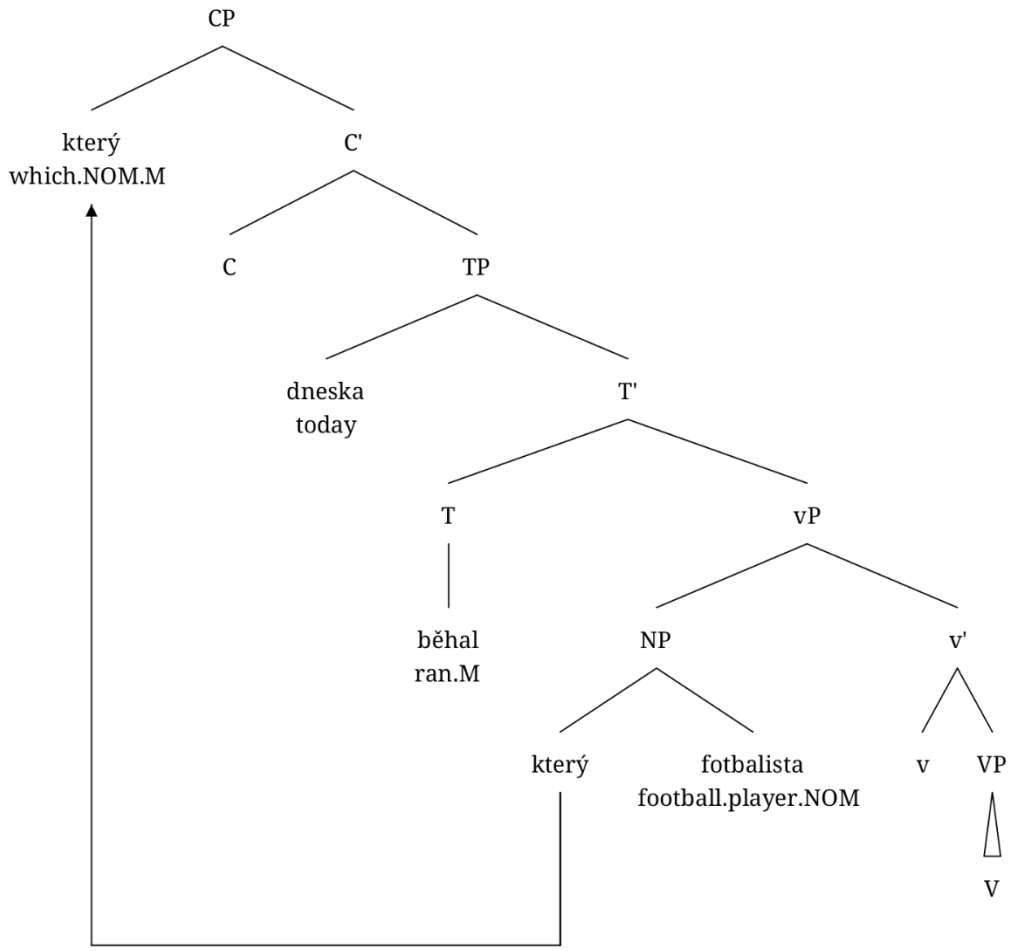
One of the variables in the fourth filler experiment is the type of adverb in the subordinate clause. Our hypothesis stands on 2 assumptions: 1) that the adverb of time is based higher in a clause than a subject, 2) that the adverb of manner is based lower than a subject. The second assumption leads to a problem with linearization and anti-locality when LBE happens from a subject.[5] There should not be an issue when an adverb of manner is present in LBE from an object, for adverbs of time (due to their higher base position) both types of extractions should be allowed and possible.

[5] The possibility of not having the anti-locality problem would be to move "který" 'which' in SpecvP to a higher position than the position of the adverb in SpecvP. This means that 'which' would have to move from SpecVP to SpecvP, which is not possible (Ko, 2007; see section 2.4.2).



Který rychle běhal fotbalista?
‘Which football player ran quickly?’

Figure 6 – visualization of a linearization problem in a structure with LBE and the adverb of manner



Který dneska běhal fotbalista?
'Which football player ran today?'

Figure 7 – visualization of LBE from a subject with intervening adverb of time

Bondarenko & Davis (2021) show the influence of the adverb type on Russian. Adapted examples are in (21). The adverb of time (high adverb) does not forbid the acceptance of LBE, the adverb of manner (low adverb) makes the extraction impossible.

(21) a Každaja$_i$ [včera večerom] [t_i devočka] vyčistila jaščik.
every.NOM.F yesterday evening.INSTR girl.NOM.F cleaned drawer
'Every girl cleaned a drawer yesterday evening.'

b *Každaja$_i$ polnostju$_k$ [t_i devočka] vyčistila jaščik t_k.
every.NOM.F completely girl.NOM.F cleaned drawer
'Every girl cleaned a drawer completely.'

The prediction for the fourth filler experiment is therefore the following: condition *a* from items 1–4 is supposed to be rated as less natural than the other three conditions.

3.3 Material

In this section, each part of the experiment will be exemplified.

3.3.1 Main experiment

Item 14 in all conditions:

a Detektiv na začátku vyšetřování tušil, kterou
detective.M at the beginning investigation.GEN suspected which.ACC.F
omáčka otrávila kuchařku.
sauce.NOM.F poisoned.F cook.ACC.F
'At the beginning of the investigation, the detective suspected which cook was poisoned by the sauce.'

b Detektiv na začátku vyšetřování tušil, která
detective.M at the beginning investigation.GEN suspected which.NOM.F
kuchařku otrávila omáčka.
cook.ACC.F poisoned.F sauce.NOM.F
'At the beginning of the investigation, the detective suspected which sauce poisoned the cook.'

c Detektiv na začátku vyšetřování tušil, kterou
detective.M at the beginning investigation.GEN suspected which.ACC.F
kuchařka otrávila omáčku.
cook.NOM.F poisoned.F sauce.ACC.F
'At the beginning of the investigation, the detective surmised which sauce did a/the cook poison.'

d Detektiv na začátku vyšetřování tušil, která
detective.M at the beginning investigation.GEN suspected which.NOM.F
omáčku otrávila kuchařka.
sauce.ACC.F poisoned.F cook.NOM.F
'At the beginning of the investigation, the detective suspected which cook poisoned the sauce.'

3.3.2. Filler experiment 1 – pronouns

Item 8 in all conditions:

a Z článku o novém projektu jsem bohužel nezjistil,
from article about new project be.AUX unfortunately did not find out
kterou ten projekt shání investorku.
which.ACC.F the project.NOM.M looks for investor.ACC.F
'From an/the article about a new project I unfortunately did not find out which investor does the project look for.'

b Z článku o novém projektu jsem bohužel nezjistil,
from article about new project be.AUX unfortunately did not find out
kterou shání investorku.

which.ACC.F (it) looks for investor.ACC.F
'From an/the article about a new project I unfortunately did not find out which investor does it look for.'

c

Z	článku	o	zdatné	investorce	jsem	bohužel	nezjistil,
from	article	about	proficient	investor.F	be.AUX	unfortunately	did not find out

který	tu investorku	shání	projekt.
which.NOM.M	the investor.ACC.F	looks for	project.NOM.M

'From an/the article about a proficient investor I unfortunately did not find out which project looks for the investor.'

d

Z	článku	o	zdatné	investorce	jsem	bohužel	nezjistil,
from	article	about	proficient	investor.F	be.AUX	unfortunately	did not find out

který	ji	shání	projekt.
which.NOM.M	she.ACC	looks for	project.NOM.M

'From an/the article about a proficient investor I unfortunately did not find out which project looks for her.'

e

Z	článku	o	zdatné	investorce	jsem	bohužel	nezjistil,
from	article	about	proficient	investor.F	be.AUX	unfortunately	did not find out

který	ta investorka	shání	projekt.
which.ACC.M	the investor.NOM.F	looks for	project.ACC.M

'From an/the article about a new project I unfortunately did not find out which project does the investor look for.'

f

Z	článku	o	zdatné	investorce	jsem	bohužel	nezjistil,
from	article	about	proficient	investor.F	be.AUX	unfortunately	did not find out

který	shání	projekt.
which.ACC.M	looks for	project.ACC.M

'From an/the article about a new project I unfortunately did not find out which project does she look for.'

g

Z	článku	o	novém	projektu	jsem	bohužel	nezjistil,
from	article	about	new	project	be.AUX	unfortunately	did not find out

která	ten projekt	shání	investorka.
which.NOM.F	the project.ACC.M	looks for	investor.NOM.F

'From an/the article about a new project I unfortunately did not find out which investor looks for the project.'

h

Z	článku	o	novém	projektu	jsem	bohužel	nezjistil,
from	article	about	new	project	be.AUX	unfortunately	did not find out

která ho shání investorka.
which.NOM.F it.ACC looks for investor.NOM.F
'From an/the article about a new project I unfortunately did not find out which investor looks for it.'

3.3.3 Filler experiment 2 – humanness balanced

Item 4 in all conditions:

a Na výstavě o vesmíru jsme zjistili, kterou
on exhibiton about space be.AUX discovered which.ACC.F
satelit fotografoval planetu.
satellite.NOM.M photographed.M planet.ACC.F
'On a space exhibition we discovered which planet did a/the satellite photograph.'

b Na výstavě o vesmíru jsme zjistili, který
on exhibiton about space be.AUX discovered which.NOM.M
planetu. fotografoval satelit.
planet.ACC.F photographed.M satellite.NOM.M
'On a space exhibition we discovered which satellite photographed a/the planet.'

c Na výstavě o vesmíru jsme zjistili, které
on exhibiton about space be.AUX discovered which.ACC.M.PL
umělkyně fotografovala astronauty.
artist.NOM.F photographed.F astronaut.ACC.F.PL
'On a space exhibition we discovered which astronauts were photographed by an/the artist.'

d Na výstavě o vesmíru jsme zjistili, která
on exhibiton about space be.AUX discovered which.ACC.M.PL
astronauty fotografovala umělkyně.
astronaut.ACC.F.PL photographed.F artist.NOM.F
'On a space exhibition we discovered which artist photographed (the) astronauts.'

3.3.4 Filler experiment 3 – definiteness of the intervening constituent

Item 4 in all conditions:

a Z mého místa za brankou nebylo vidět, kterého
from my seat behind goal be.AUX.NEG see.INF which.ACC.M
kapitán porazil rozhodčího.
captain.NOM.M knocked down referee.ACC.M
'From my seat behind the goal I could not see which referee did the captain knock down.'

b Z mého místa za brankou nebylo vidět, který
from my seat behind goal be.AUX.NEG see.INF which.NOM.M
kapitána porazil rozhodčí.

captain.ACC.M knocked down referee.NOM.M
'From my seat behind the goal I could not see which referee knocked down the captain.'

c Z mého místa za brankou nebylo vidět, kterého
from my seat behind goal be.AUX.NEG see.INF which.ACC.M
hráč porazil rozhodčího.
player.NOM.M knocked down referee.ACC.M
'From my seat behind the goal I could not see which referee did a player knock down.'

d Z mého místa za brankou nebylo vidět, který
from my seat behind goal be.AUX.NEG see.INF which.NOM.M
hráče porazil rozhodčího.
player.ACC.M knocked down referee.NOM.M
'From my seat behind the goal I could not see which referee knocked down a player.'

3.3.5 Filler experiment 4 – adverbs

Item 2 (with the extraction from a subject) in both conditions:

a Studenti se dohadovali, který nudně mluvil přednášející.
students argued which.NOM.M boringly talked.M lecturer.NOM.M
'(The) Students argued which lecturer talked boringly.'

b Studenti se dohadovali, který včera mluvil přednášející.
students argued which.NOM.M yesterday talked.M lecturer.NOM.M
'(The) Students argued which lecturer talked yesterday.'

Item 7 (with the extraction from an object) in both conditions:

a Knihkupec se zákaznice přeptával, kterou těžce
bookseller REFL customer.F asked which.ACC.F hardily
sháněla knihu
looked for.F book.ACC.F
'The bookseller asked a/the customer which book did (she) look for hardily.'

b Knihkupec se zákaznice přeptával, kterou předevčírem
bookseller REFL customer.F asked which.ACC.F day before yesterday
sháněla knihu
looked for.F book.ACC.F
'The bookseller asked a/the customer which book did (she) look for the day before yesterday.'

3.3.6 Filler experiment 5 – control

Item 1 as a representant of supposed nonproblematic and natural clauses:

Herec se domlouval s rekvizitářem kterou vázu

actor	REFL	agreed on.M	with	props man	which. ACC.F	vase.ACC.F
ve	scéně	rozbije.				
in	scene.DAT.F	will break				

‘An/the actor agreed with the props man on which vase (he) will break in the scene.’

Item 18 as a representant of supposed nonnatural clauses:

Šéfkuchař	se	dohadoval	s	hostitelem	které	maso
chef	REFL	argued.M	with	host	which.ACC.N	meat.ACC.N
zítra	mu	uvaří.				
tomorrow	he.DAT	will cook				

‘The chef argued with the host about which meat (he) will cook him tomorrow.’

3.3 Participants

The participants were obtained via the seminar “Participation in linguistic and psychological experiments in LABELS Lab” with the help of doc. Mgr. Jan Chromý, Ph.D. The university students (who were from different faculties, not only students of philology) filled out the experiment in return for credits. After being sent the link to the experiment, the participants had one week to fill it in. Altogether, 142 participants completed the experiment.

3.3.1 Procedure

The whole experiment was conducted on the L-Rex platform. The participants were instructed to rate the naturalness of sentences (one at a time) on a scale from 1 to 7, where 1 was “not natural” and 7 was “natural”. Each participant was shown 80 items (for every item just one condition) and one clause at a time. The items were pseudo-randomized.

3.3.2 Filtration

After collecting all the data, the results were filtered based on time spent on filling in the whole experiment and on ratings from filler experiment 5 (as mentioned in 3.1.6). The boundary for time was set on 500 seconds, time lower than this would give the participants approximately 6 seconds for each item. We chose (prior to the run of the experiment) 8 items from the fifth filler experiment to function as a measure of ‘reliability’ of the participants. Half of these items had a standard word order and thus should be rated with a higher number on the scale (items 1, 5, 7, 11). The other half contained of supposed agrammatical items (with nonstandard word order), and the ratings should therefore be lower on the scale (items 13, 18, 22, 24). For the first mentioned, we set the limit for good rating on 5 and more (everything less than 5 was filtered). For the latter, we decided to draw the borderline on 3,5, the participants with ratings lower than that were left. This was taken in consideration with the difference between the “good” and the “bad” items, if the difference between the mean ratings of both was smaller than 2, these

participants were filtered. Lastly, the participants whose "good" and "bad" ratings were reversed were eliminated from the data.

In the end, we were left with 70 participants whose data were then analysed.

3.4 Results

The results are presented with boxplots to indicate the range of ratings. The figures with z-scores were added to better show the tendencies of acceptability of the conditions.

The first thing apparent from the results is that the clauses with LBE are generally perceived as unnatural, throughout the experiment the majority of participants rated the constructions on the lower portion of the scale.

3.4.1 Main experiment

As is evident from figures 4 and 5, only one condition shows a difference from the others. It is condition *c* where a human subject intervenes the extraction from a nonhuman object. Conditions *a*, *b* and *d* have close similarities as for their rating.

The constituent order alone in the main experiment does not influence the acceptability as well as the order of humanness. The interaction between human order and constituent order is statistically significant.

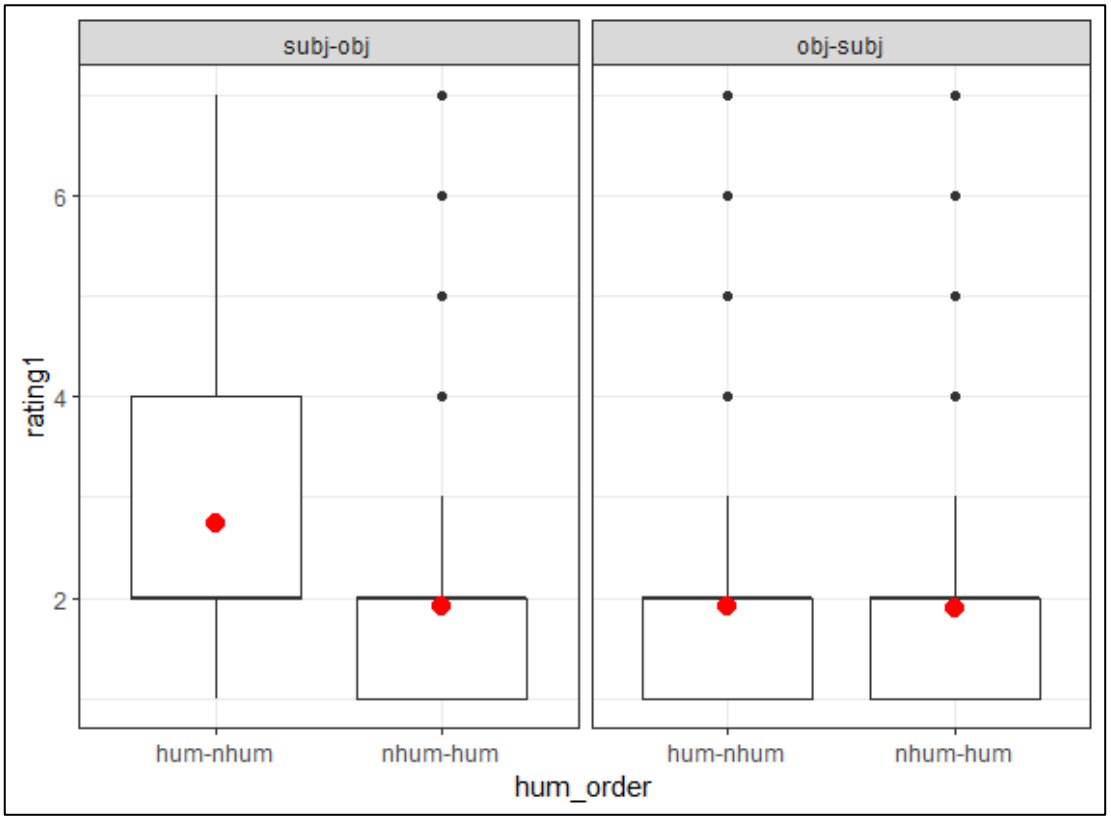


Figure 8 – Main experiment: boxplot

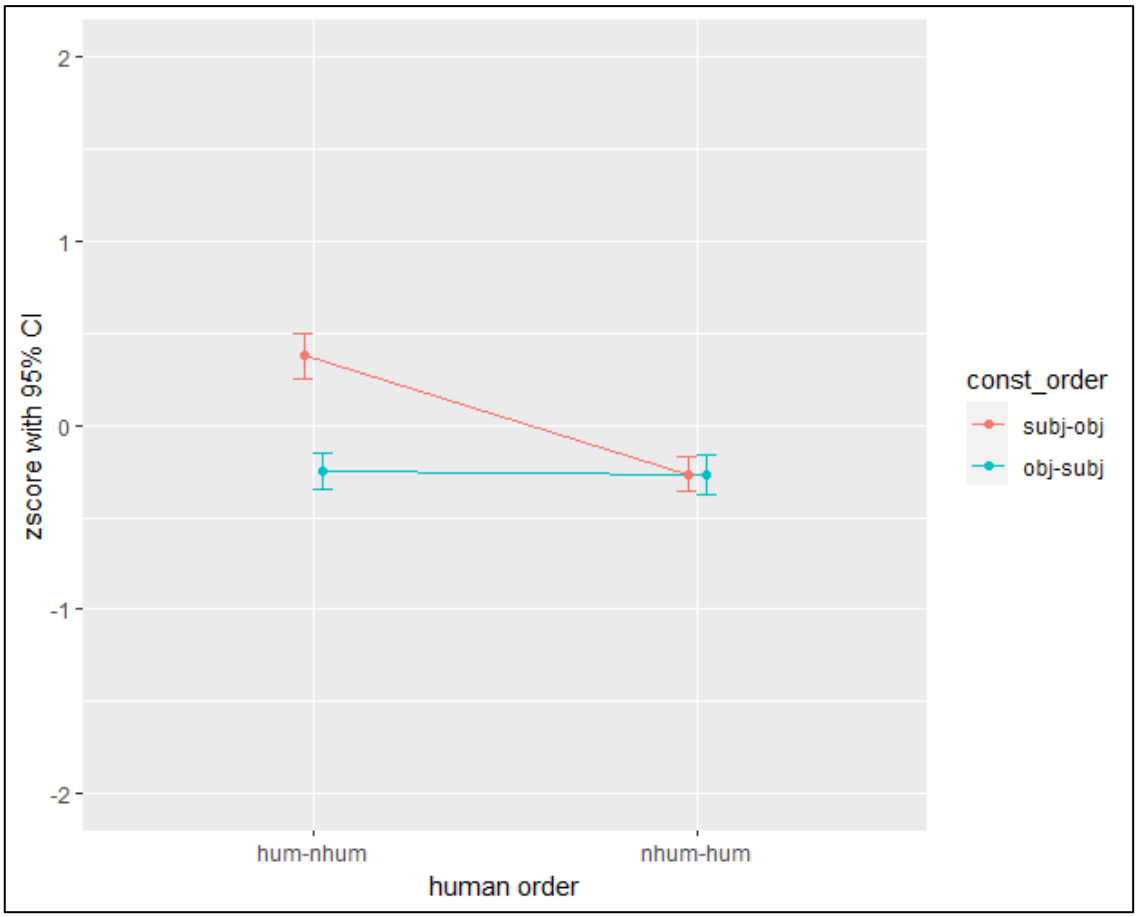


Figure 9 – Main experiment: z-scores with 95% CI

3.4.2 Filler experiment 1 – pronouns

The items in filler experiment 1 were rated as one of the most natural clauses in the experiment. The conditions with constituent order subject – object (*a*, *b*, *e,* and *f*) have higher ratings than the order object – subject. The nature of the intervening constituent reflects on the naturalness of the clause in a way that the overt NP is worse than the covert (or omitted) constituents. In the case of the extraction from an object (constituent order subject – object), the order human – nonhuman was rated as slightly better than the opposite order. Interestingly, the participants stated that the extraction from a subject is more acceptable if the intervening constituent in an overt form is nonhuman. Condition *c* has the lowest rating from this filler experiment.

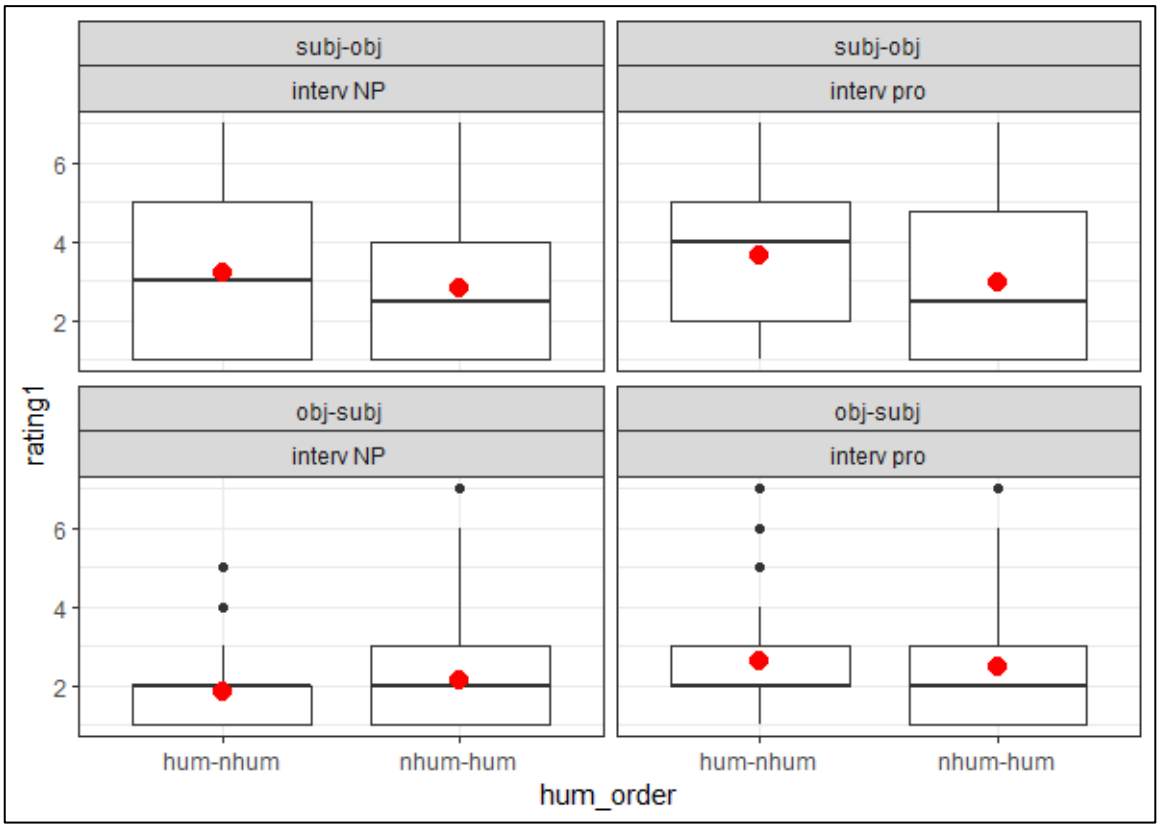


Figure 10 – Filler experiment 1: boxplot

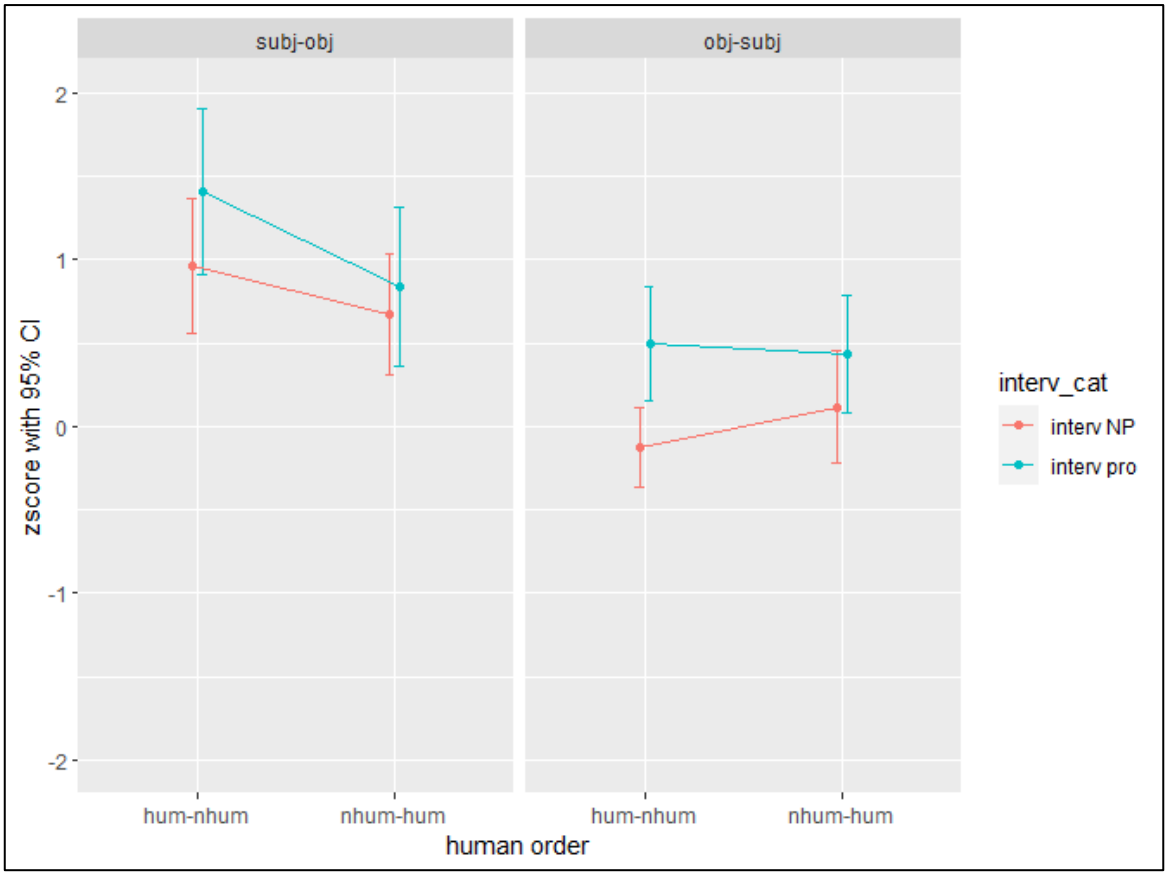


Figure 11 – Filler experiment 1: z-scores with 95% CI

3.4.3 Filler experiment 2 – humanness balanced

In the second filler experiment, the type of extraction reflects strongly on the naturalness of the clauses. Both conditions with the subject – object order are rated higher than the ones with the opposite constituent order. As for humanness, there is only a slight difference in the acceptability, the clauses with nonhumans have lower ratings than the ones with only humans.

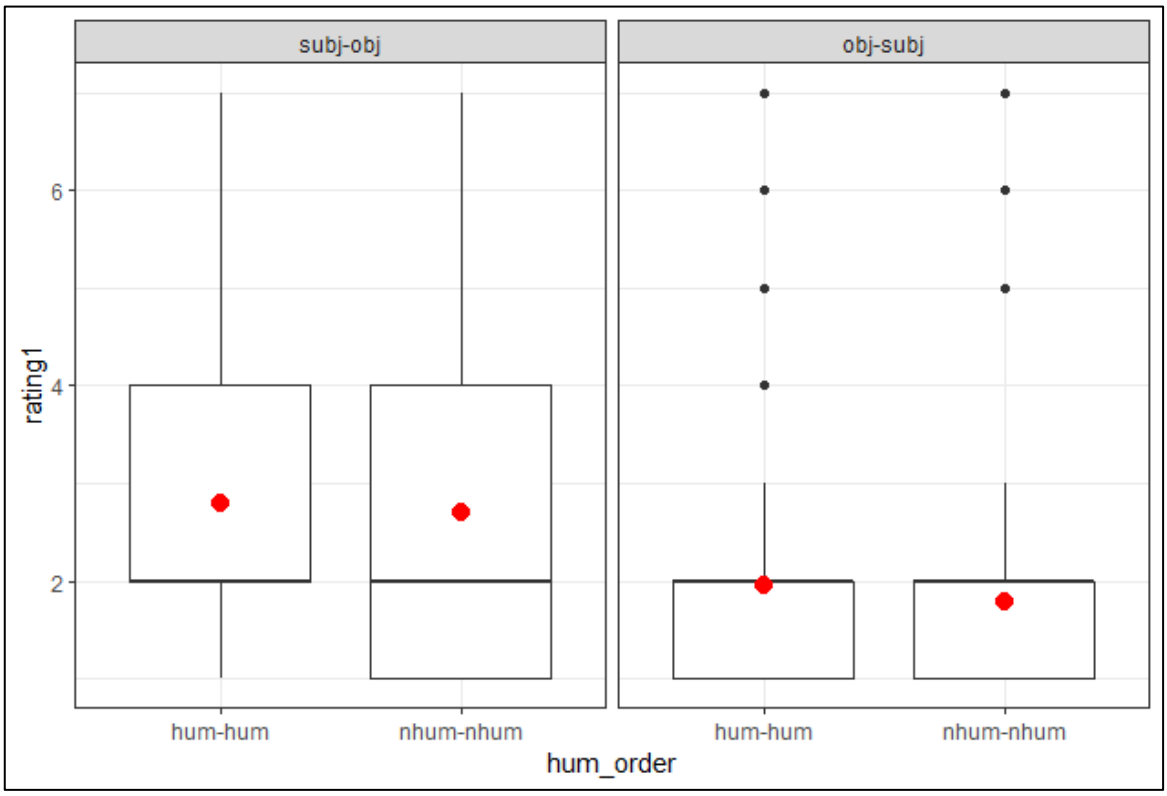


Figure 12 – Filler experiment 2: boxplot

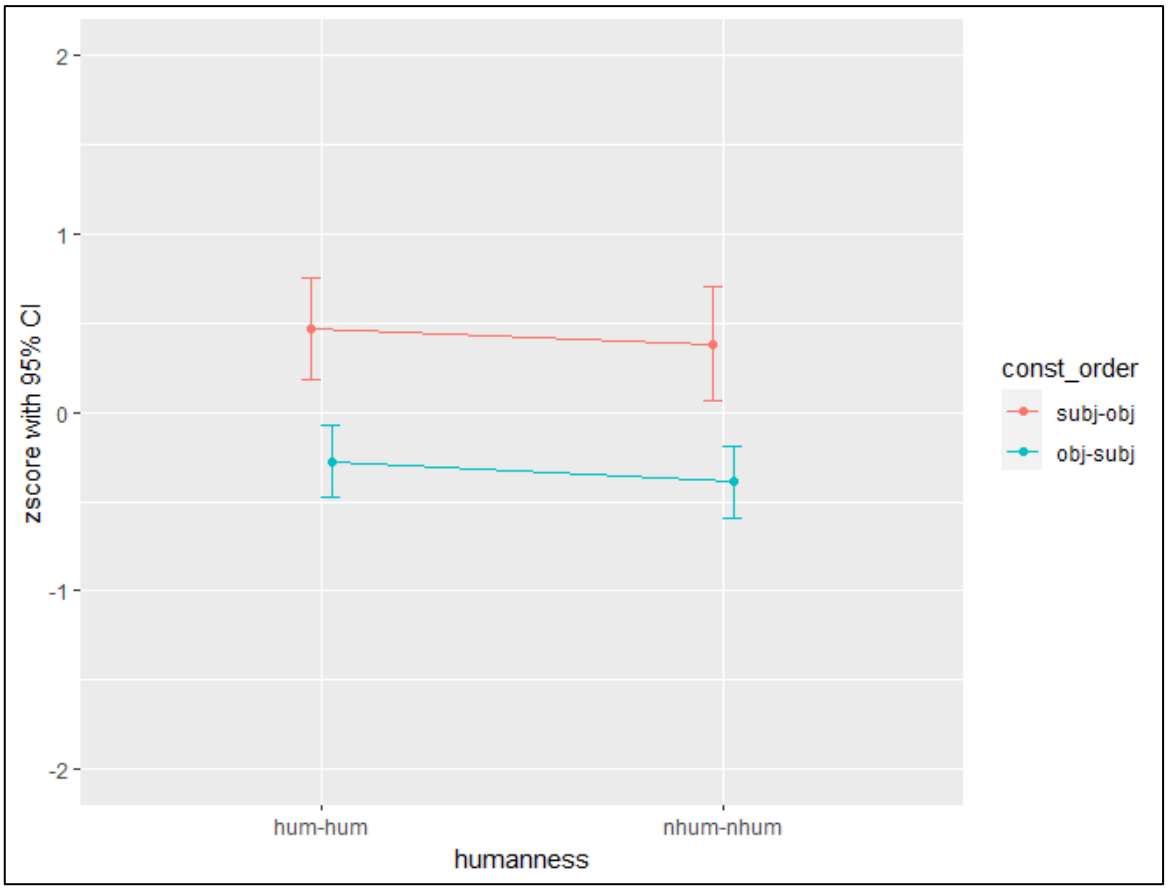


Figure 13 – Filler experiment 2: z-scores with 95% CI

3.4.4 Filler experiment 3 – definiteness of the intervening constituent

The results of the third filler experiment again show the same higher rating of clauses with the constituent order subject – object. In both types of extraction, the intervening constituent being nondefinite leads to higher ratings (in the extraction from a subject only slightly).

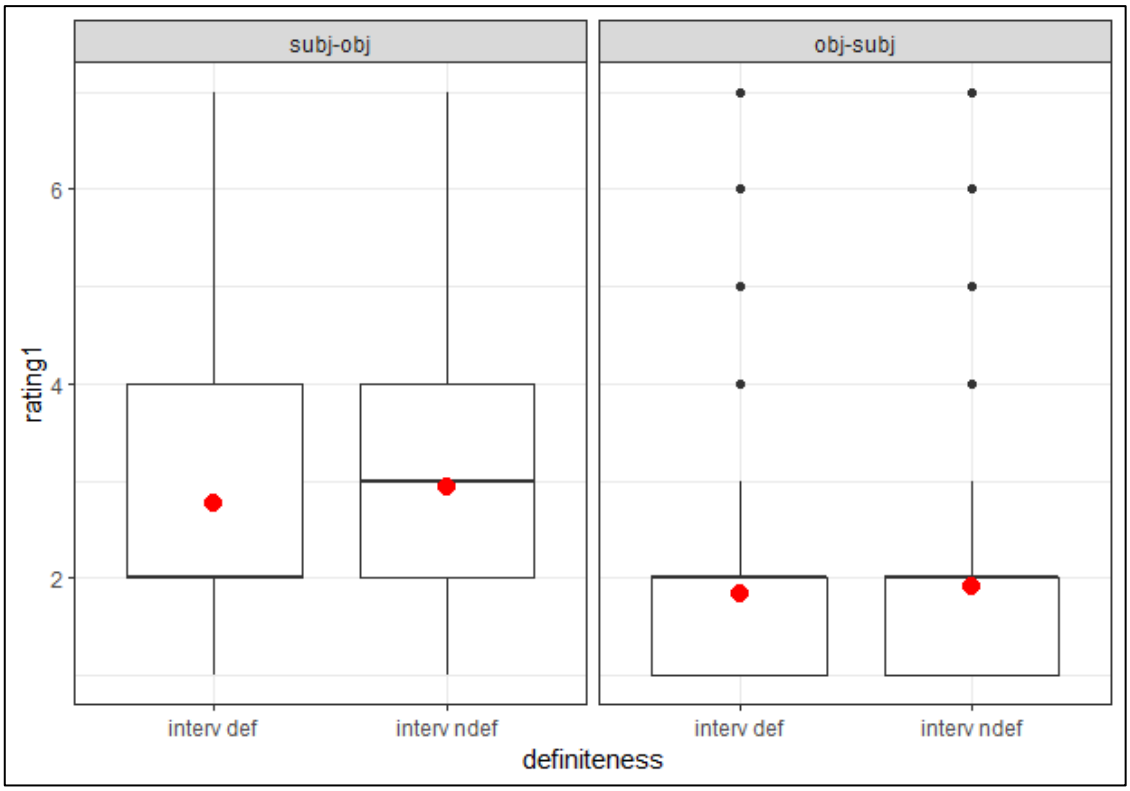


Figure 14 – Filler experiment 3: boxplot

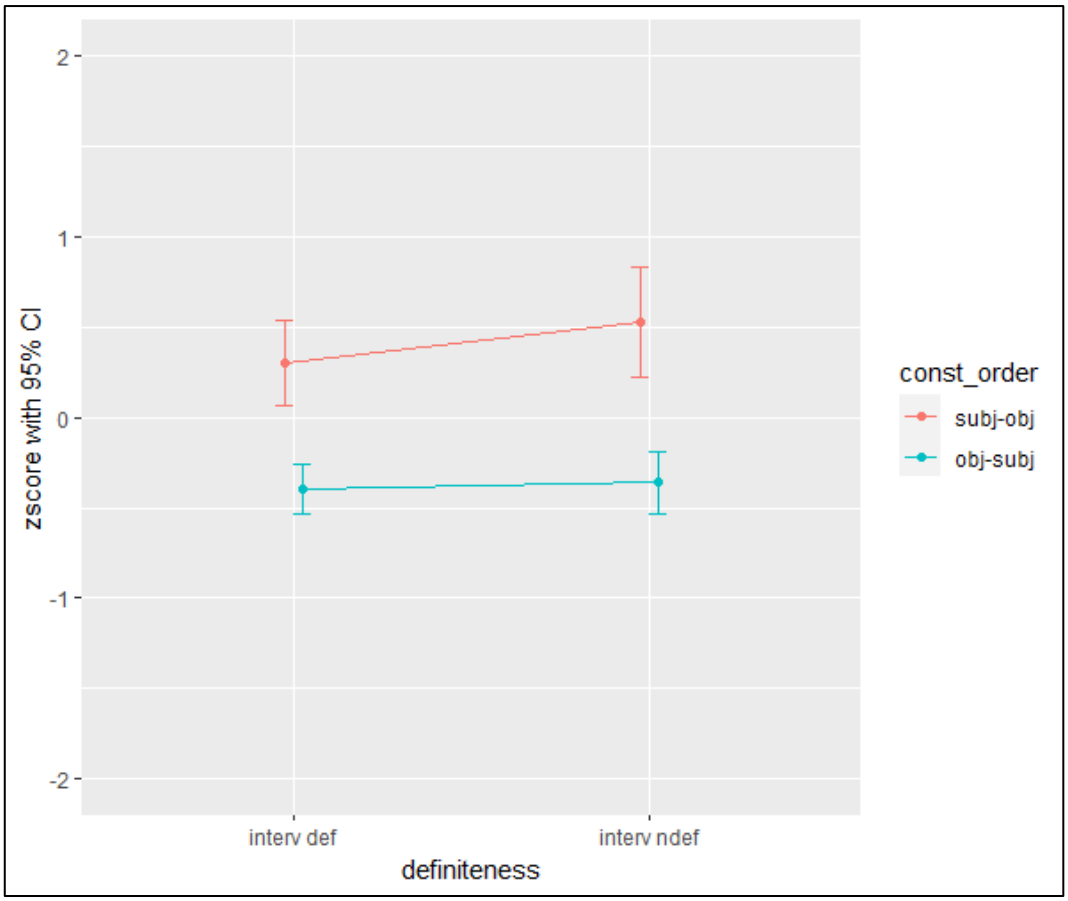


Figure 15 – Filler experiment 3: z-scores with 95% CI

3.4.5 Filler experiment 4 – adverbs

The adverb type of time in clauses is rated higher than the items with an adverb of manner. The best condition is then the one with the extraction from an object paired with an adverb of time. Altogether, both variables were significant.

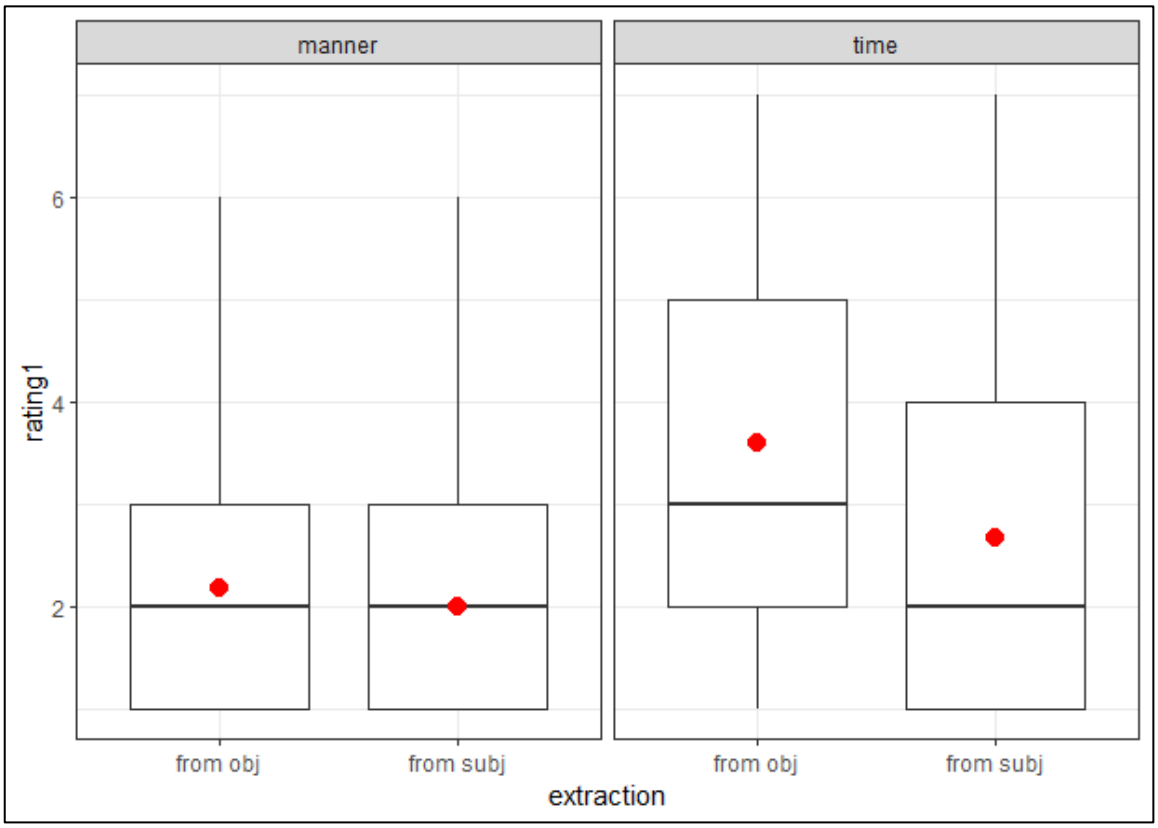


Figure 16 – Filler experiment 4: boxplot

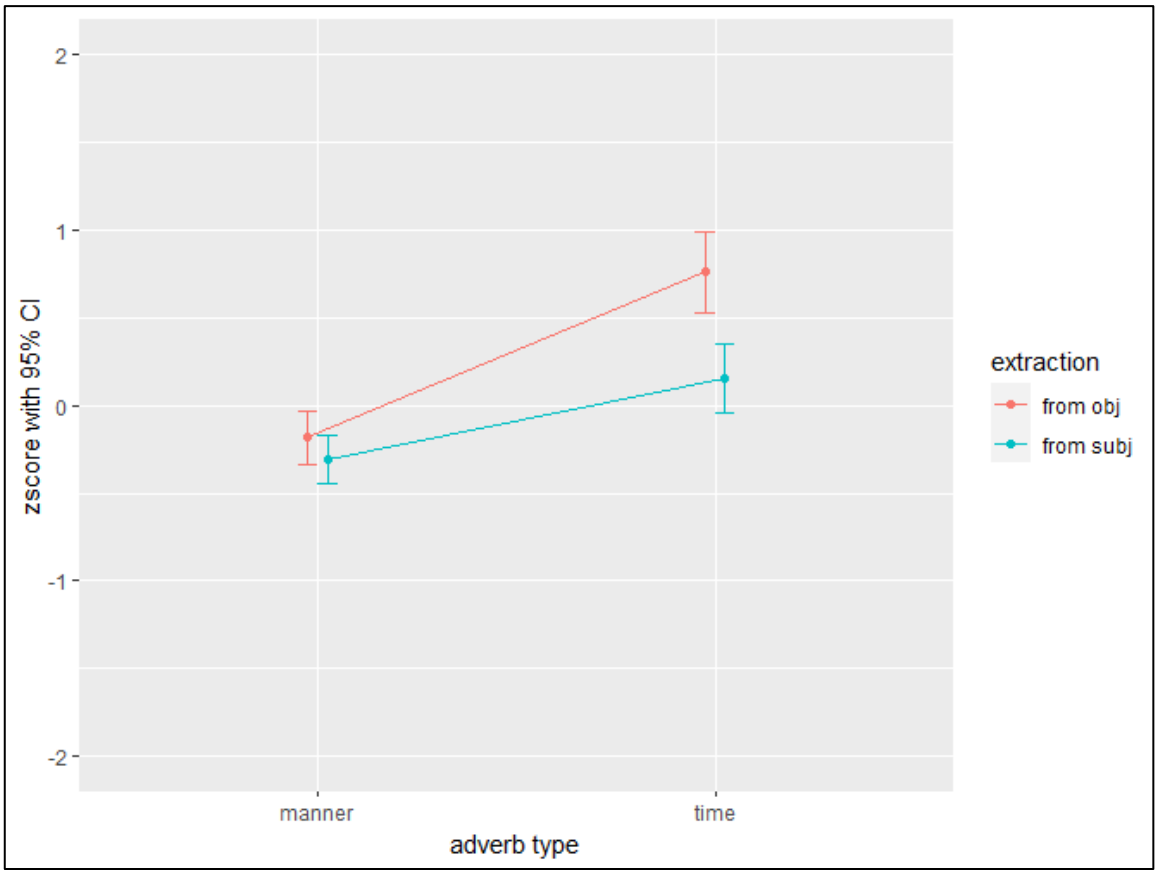


Figure 17 – Filler experiment 4: z-scores with 95% CI

3.5 Discussion

Our first question was whether the word order in Czech is strictly based on the syntactic function of the constituents or if the humanness of both subject and object also decide on the acceptability of the clauses. Our predictions based solely on the subject – object word order were not fulfilled in the main and the first filler experiment.

Deriving from this, the humanness of the constituents indeed affects the word order. The low rating of condition *b* in the main experiment rules out the analysis (in the spirit of Fanselow, 2001) using the relaxed UTAH. However, the analysis (in the spirit of Fanselow, 2003) reflecting the UTAH and using the binding of a semantic variable meets the results. Our suggested canonical order is the only possible for the acceptance of clauses with LBE. The results of the main experiment surely state, that this prediction was correct. All conditions where the word order was not subject human – object nonhuman were rated similarly badly. In the first filler experiment, the two conditions (*e* and *f*) where the word order was a nonhuman object following a human subject were also rated as the most natural in the subexperiment.

As far as the word order in clauses with balanced humanness goes, our prediction for the acceptable word order being subject – object was also fulfilled. In both the second and the third

filler experiments, the clauses with the extraction from an object (and the intervening constituent on the first place thus a subject) were rated higher than the ones with the extraction from a subject.

The definiteness of the intervening constituent in filler experiment 3 plays with the naturalness of the clauses in the opposite way than we predicted. The nondefinite intervening constituents indicate a better acceptability of the whole clause.

As mentioned in section 3.4.2, the items in F1 were the most natural from the entire experiment. This is most certainly due to the fact that the intervening constituent in each clause was also included in the main clause. The contextual involvement thus leads to a better acceptance of clauses.

Lastly, the results of the fourth filler experiment are quite puzzling. Our prediction was that only the adverb of manner as an intervening constituent of the extraction from a subject would be rated worse than the others. The clauses with an adverb of manner were, however, rated poorly in both types of extractions.

Altogether, the results confirm the important aspect of the semantic nature of the constituents. Basic word order in Czech is not strictly based on the syntactic function of its constituents if the humanness of the two (more specifically of the subject and the object) is different. For these cases, the acceptable word order is a subject that is a human followed by an object that is nonhuman. When the humanness of both constituents is the same, the word order reflects the syntactic function, thus the word order is subject – object.

4 Conclusions

Left branch extraction is a fascinating syntactic phenomenon. Its character of being a movement out of an NP (in our case, in Czech) can show us some insights into the basic word order due to the fact that this kind of movement is limited. Through many examples we came to a direction of the possibilities of what is permitted and what not.

After getting introduced to the phenomenon by itself, we discovered some similarities of LBE in Czech with other Slavic languages, mostly Russian and Polish. The first restrain on the extraction is in structures where NP is a complement of a PP. In larger left branches of NPs plays to the acceptance of a whole clause the order of constituents in the left branch itself. Probably the most relevant observation (and also the one that stood at the beginning of this thesis) is the intervention of other constituents to the extraction. The extraction from a subject over an object seems to cause unacceptability of the clause but the extraction from an object over a subject is fine. After looking at the animacy of constituents and its effect on the word order, two hypotheses of analysis were introduced. To be able to form a clause where a human object is on a prior position to the nonhuman subject, the constituents can be base-generated in their opposite position, i.e., the object is base-generated on the canonical base position of subject and vice versa (and thus relaxing the UTAH). The other option relies on the human object to be base-generated in SpecvP as well as the subject, the latter being lower. The object than binds a semantic variable on its canonical base position. In conclusion, we came to proposal of the basic Czech word order to be the following: the prior position is occupied by a subject that is a human, an object that is semantically nonhuman is on the latter position. If the animacy status is the same for both subject and object, the word order is subject – object.

The problem of unacceptability of some clauses with LBE is not solely the type of extraction, rather, it is the need of scrambling for the creation of the desired word order. If the intervening constituent is to be scrambled, the clause should not be possible. The cyclic linearization theory explains the movement of a constituent out of a phase without going through its edge with the preservation of word order after the linearization of vP and after the linearization of CP. In other words, if no scrambling happens, the structure is possible.

The empirical part of this thesis showed that the theoretical claims previously constructed for Russian material mostly work for Czech, too. Even though, the items containing LBE were mostly rated quite low, the main prediction that the clauses with LBE and without scrambling (i.e., the clauses which followed the word order subject human – object nonhuman / subject – object for balanced humanness) was indeed confirmed in the results. Furthermore, based on the

results, the analysis suitable for explaining the word order being a human preceding a nonhuman reflects the UTAH and puts both the nonhuman subject and the human object in a SpecvP position. The object in a higher position then binds a semantic variable in its canonical position (as a complement of V)

Yet, there are still other aspects of structures with LBE (and Czech basic word order in general) which are to be tested and looked upon. The filler experiments showed some tendencies for the overtness and the definiteness of constituents, as well as a possible intervention of high and low adverbs. These can work as basis for future studies. In addition, LBE can further be studied in relation to spoken language or from the point of pragmatics. Another direction for further research can be found in testing the difference of the effect on the word order between a human and an animal (or a thing and an animal).

All in all, the analysis consisting of the phase theory in combination with cyclic linearization proposed by Bondarenko & Davis (2021) proved to be a valid analysis supported by empirical evidence from the presented experiment.

5 References

BAKER, M. C.: *Incorporation: A Theory of Grammatical Function Changing*. Chicago, Illinois: The University of Chicago Press, 1988.

BISKUP, P.: Left branch condition. In: Karlík, P.; Nekula, M; Pleskalová, J. (eds.), *CzechEncy – Nový encyklopedický slovník češtiny* [online].

BISKUP, P.: Podmínka fázové impenetrability. 2017 In: Karlík, P.; Nekula, M; Pleskalová, J. (eds.), *CzechEncy – Nový encyklopedický slovník češtiny* [online].

BONDARENKO, T.; DAVIS, C.: Sub-extraction asymmetries and linearization in Russian. *Journal of Slavic Linguistics*. 2021, 29 (FASL 28 extra issue).

BOŠKOVIĆ, Ž.: What will you have, DP or NP? In Elfner, E.; Walkow, M (eds.), *Proceedings of NELS* 2008, 37, 101–115.

CORVER, N. F. M.: *The syntax of left branch extractions*. Dissertation. Tilburg University, 1990.

FANSELOW, G.: Der Rückkehr der Basisgenerierer. *Groninger Arbeiten zur germanistischen Linguistik,* 1993, 36, 1–74.

FANSELOW, G.: Free Constituent Order: A Minimalist Interface Account. *Folia Linguistica*. 2003, 37 (1–2), 191–232.

FANSELOW, G.; LENERTOVÁ, D.: Left peripheral focus: mismatches between syntax and information structure. *Natural Language & Linguistic Theory*. 2011, 29(1), 169–209.

FOX, D.; PESETSKY, D.: Cyclic linearization of syntactic structure. *Theoretical Linguistics*. 2005, 31 (1–2), 1–45.

HAJIČOVÁ, E.; HAVELKA, J.; SGALL, P.; VESELÁ, K.; ZEMAN, D.: Issues of projectivity in the Prague Dependency Treebank. *The Prague Bulletin of Mathematical Linguistics.* 2004, 81, 5–22.

CHAVES, R. P.: Island phenomena and related matters. In: Müller, S.; Abeillé, A.; Borlsey, R.D.; Koenig, J.-P. (eds.), *Head-Driven Phrase Structure Grammar: The handbook*. Berlin: Language Science Press, 2021, 665–725.

CHOMSKY, N.: Minimalist inquiries: The framework. In Martin, R.; Michaels, D.; Uriagereka, J. (eds.), *Step by step: Essays on minimalism in honor of Howard Lasnik*. Cambridge, MA: MIT Press, 2000.

JASINSKAJA, K.; ŠIMÍK, R.: Slavonic free word order. In Fellerer, J.; Bermel, N. (eds.), *The Oxford guide to the Slavonic languages*. Oxford: Oxford University Press, in print.

KO, H.: Asymmetries in scrambling and cyclic linearization. *Linguistic Inquiry* 38(1). 2007, 49–83.

NEELEMAN, A.: Scrambling as a D-structure phenomenon. In: Corver, N. & van Riemsdijk, H. (eds.), *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*. Berlin: Mouton de Gruyter, 1994, 387–429.

ROSS, J. R.: *Constraints on Variables in Syntax*. Dissertation. MIT, 1967.

SGALL, P.; HAJIČOVÁ, E.; BURÁŇOVÁ, E.: *Aktuální členění věty v češtině*. Praha: Academia, 1980.

TITOV, E.: *Information Structure of Argument Order Alternations*. Dissertation. University College London, 2012.

UHLÍŘOVÁ, L.; KUČEROVÁ, I.: Slovosled. In: Karlík, P.; Nekula, M; Pleskalová, J. (eds.), *CzechEncy – Nový encyklopedický slovník češtiny* [online].

VESELOVSKÁ, L.: *Wh-questions: a case study in Czech*. Olomouc: Palacký University Olomouc, 2021.