sciendo
DOI 10.2478/jazcas-2024-0045

CLITICS IN SLOVAK WITH(OUT) CONSTRAINTS. CORPUS FREQUENCY VERSUS ACCEPTABILITY RATINGS

MARTINA IVANOVÁ
Institute of Slovak and Media Studies, Faculty of Arts, University of Prešov, Slovakia
Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences, Bratislava, Slovakia

IVANOVÁ, Martina: Clitics in Slovak with(out) constraints. Corpus frequency versus acceptability ratings. Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 291 – 308.

Abstract: Using data from a representative corpus of Slovak and an acceptability survey, the preferential placement of clitic components in object clause constructions in Slovak have been investigated. Slovak clitics are usually described as elements following Wackernagel's Law and belonging to the category of second-position clitics. However, usage-based investigations show that their placement varies within a clause, depending on various pragmatic and syntactic factors and a set of constraints which limit their movement within the clause structure. By comparing data from corpus analysis and acceptability ratings by native speakers, it is shown how judgments and actual usage of clitics may converge or diverge in particular cases.

Keywords: clitics, word order, Wackernagel's Law, corpus, acceptability judgements, barriers, clitic climbing, object clause constructions.

1. INTRODUCTION

In most languages, sentence constituents may be linearized in two or more different ways, at least in some well-formed sentences. Nevertheless, all natural languages are restrictive in one way or another: no language allows for all possible linear orders of sentence categories in 100% of sentences, linearization constraints are salient for all word order systems. In Slavic languages like Slovak, most combinations of scrambling types are available for sentence categories represented by non-clitic words, while the number of scrambling types available for clitics is more reduced (cf. Zimmerling 2011, p. 754). The goal of the paper is to investigate possible patterns of clitic placement in object control clauses on the basis of the corpus data and to answer the question how these word order variants are evaluated by speakers in an acceptability rating experiment.

Clitics represent one of the most specific and intricate word order phenomena of many languages. Slovak belongs to those languages which follow Wackernagel's Law, and its clitic elements belong to the category of second-position clitics (2P) (cf. Franks – Holloway King 2000). However, Wackernagel's Law is not always

applicable without exception and the pattern of clitic placement undergoes systematic variations under certain conditions. The cues from the grammar system and information structure can interfere in the linearization patterns of clitic components and remove them from the second position.

The paper is structured as follows: The introductory remarks are presented in Section 1. Section 2 considers phonological and non-phonological definitions of clitics as usually presented in various linguistic approaches. Section 3 provides a concise overview of the methodological framework adopted in the paper as well as the design of the dataset including both corpus and experimental data. In Section 4, results of the corpus investigation and survey of clitics placement in object control clauses are presented. Section 5 summarizes the findings.

2. CLITICS IN SLOVAK

2.1. Definition of clitics

Clitics can be compared with full words and affixes.

As opposed to full words, clitics are typical of:

(i) prosodic deficiency: they are elements "that lack independent stress" (Pescarini 2021, §1.1), they are words in the morpho-syntactic sense, but not in the phonological sense (Booij 2012, p. 290), therefore they are unable to appear sentence-initially;

(ii) bondedness: they cannot occur in isolation, they are "defective in their phonological representation and therefore have to prosodically combine with an adjacent non-clitic word" (Ionova 2019, p. 22), usually termed as prosodic host.

In their paper on the English negative *-n't*, Zwicky and Pullum (1983, pp. 502–504) contrast clitics with affixes. There are two characteristics which set clitics apart from affixes:

(i) non-selectivity or promiscuity: they are typical of the lack of word-class selectivity, i.e. they are not selective with regard to their host;

(ii) morphological stability: affixed words tend to display morphophonological and semantic idiosyncrasies; clitic groups do not (cf. Zwicky – Pullum 1983, p. 504).

Clitics can be described as elements with "triple" citizenship. Phonologically, they lean on their prosodic hosts; positionally, they precede or follow their structural host or anchors; and functionally, they form morphological, lexical, or syntactic units with their matrix item. Depending on their contextual environment, the roles of prosodic host, anchor, and matrix item can overlap and be expressed by the same sentence component or, alternatively, different sentence components can fulfil the roles of prosodic host, anchor, and matrix item.

According to Haspelmath (2023) anchor is the word preceding an enclitic, and the word following a proclitic, whereas host is the element with which a clitic forms a prosodic word.

In example (1), *by* is an enclitic as it cannot occur at the beginning of a free form (**by bolo*), the infinitive *prehovárať* can be described as an anchor as it precedes the enclitic *by* and the component *by* forms a prosodic word with verb form *bolo* so that *bolo* is its prosodic host. At the same time, *bolo by* is an analytical grammatical form in Slovak, so that *bolo* is a matrix expression for the conditional component *by.*

(1) *Prehovárať ju* ***by*** *bolo zbytočné.*
persuade-INF her-ACC COND be-PAST-NEUTR.SG useless
'It would be useless to persuade her.'

However, in word order variant (1a), the verb *bolo* fulfils both the role of anchor and host.

(1a) *Bolo* ***by*** *zbytočné prehovárať ju.*
be-PAST-NEUTR.SG COND useless persuade-INF her-ACC

It confirms the claim of Franks and Holloway King (2000) who argue that the direction of prosodic attachment of clitics is underspecified, i.e. that it can attach to a host to their right as well as to a host to their left.

2.2. The sets of investigated clitics

The following set of criteria can be applied in classification of clitic components:

(i) Tenacity criterion: refers to the fact whether clitic items keep their clitic status in different contextual environments:

- clitics tantum (or constant 'clitics' – cf. Rosen 2001, Hana 2007; pure clitics – cf. Avgustinova – Oliva 1997) are elements which always appear in the second position;

- volatile clitics (or semi-clitics – cf. Avgustinova – Oliva 1997) can have phonological autonomy under certain contextual conditions and thus appear in the second position only optionally.

The clitic status of semi-clitics can be proved by their realization within the clitic cluster. On the basis of the rule described in Hana (2007, p. 76), element X between 1P and clitic component is a clitic:

(2) *Oni* ***nám*** ***ho*** *vzali.*
They-NOM us-DAT it-ACC take-PAST-PL
'They took it from us.'

As the semi-clitic component *nám* is interposed between 1P (*Oni*) and the permanent clitic (*ho*), it can be considered a clitic;

(ii) Functional criterion: refers to the functional status of the clitic component:

(a) auxiliary verbal clitics: assist main verbs in conveying person and number grammatical meanings in the past participle forms: *(písal) som* '(wrote)-PRES.SG.1', *si* '(wrote)-PRES.SG.2', *sme* '(wrote)-PRES.PL.1', *ste* '(wrote)-PRES.PL.2';

(b) non-reflexive argument clitics: clitics that refer to arguments of the verb:

- short forms of personal pronouns with existing long counterparts: *ma* 'me-ACC' (as opposed to *mňa*), *ťa* 'you-SG.ACC' (as opposed to *teba*), *ho* 'him-ACC' (as opposed to *jeho*, *neho*), *mi* 'me-DAT' (as opposed to *mne*), *ti* 'you-SG.DAT' (as opposed to *tebe*), *mu* 'him-DAT' (as opposed to *jemu*, *nemu*),

- short forms of personal pronouns missing long counterparts: *nás* 'us-ACC', *vás* 'you-PL.ACC', *ich* 'them-ACC', *ju* 'her-ACC', *jej* 'her-DAT', *nám* 'us-DAT', *vám* 'you-PL.DAT', *im* 'them-DAT';

(c) reflexive clitics *sa*, *si*: are notoriously ambiguous when it comes to their functional status, usually described as verbal components or pronominal expressions;

(d) particle clitics: a special form of the originally auxiliary *be*-form *by* used to build periphrastic form of the conditional mood.

3. METHODOLOGY AND DESIGN OF DATASETS

3.1. Aim and framework

The present paper combines synchronic corpus analyses with an experimental method, namely acceptability judgements to assess frequency distributions and speakers' acceptance of different word-order variants concerning clitic placement in Slovak. This approach stems from those works which emphasize that grammaticality can be operationalized by acceptability, e.g. Riemer (2009).

Acceptability judgments are found by many researchers to be a useful source of data, although, as with any source, they must be used carefully (see e.g. Sprouse 2007, 2008, 2009) for a discussion of the limits of acceptability judgments.

Research experiments on possible correlations between corpus data and acceptability ratings have revealed so far that there is a correlation between corpus data and acceptability, but it is not proportional or symmetric, i.e. we cannot count on the proportions to correspond precisely to value judgments, nor is it always possible to abstract predictions about acceptability from corpus data (Bermel – Knittl 2012, p. 246). Kempen and Harbusch (2005) as well as Bader and Häussler (2010) find considerable support for the thesis that corpus frequency is not a fully reliable predictor of acceptability. This phenomenon, known as the "frequency/acceptability mismatch", also called the "grammaticality/frequency gap", refers to the observation that there is no reliable correlation between the frequency of a grammatical unit and its acceptability.

In the following parts, the present paper brings results of a corpus and survey analysis which were conducted to understand the factors determining the word order variability of clitic components.

3.2. Dataset design

The data for the current research are of twofold sources: corpus data and experiment based on an acceptability judgment task.

Corpus data for this research were retrieved from corpus Omnia Slovaca IV Maior Beta which has 6 596 573 997 tokens, and it is compiled by the Slovak National Corpus and web corpora. Occurrences of word order pattern are retrieved from the corpus using CQL queries in which morphological tags with word- and lemma-based attribute searches are combined. To get information on variation in the word order patterns, first individual word order variants were searched for, and then sifted manually to remove erroneous results.

Experimental data were summoned using a survey which took place in spring 2023 over the web in the form of an online questionnaire. All in all, 153 respondents (the majority of them were students at different universities or teachers at schools across the country who were recruited online) were asked to evaluate different word order variants of the same structures with respect to their acceptability. Table 1 brings a closer look at the characteristics of the respondents' sample.

Table 1. Age, profession and gender of respondents

Age	n =	Profession	n =	Gender	n =
0 – 18	2	Linguists	36	M	33
19 – 35	23	Non-linguists	117	F	120
26 – 35	40				
36 – 45	35				
46 – 55	27				
56 – 65	10				
66 – 75	6				

To assess the acceptability of certain word order patterns, gradient Likert's scale with numerical values from 1 to 5, i.e. from fully acceptable vs. fully unacceptable was used. The scale had descriptors at all points from 1 to 5. For the individual stimuli, examples were taken from the corpus wherever possible, sometimes simplifying and toning them down to avoid having respondents react to irrelevant elements in the sentence. The order of word order patterns was randomized.

Responses on the Likert scale are regarded as ordinal rather than interval data, suggesting that non-parametric tests should be our first resort. However, properly designed and implemented Likert-scale linguistic surveys are often subjected to parametric analysis, which can, in many instances, be more accurate and revealing. Correlation tests, which are commonly used on experimental data to show relationships between the variables, were used for the analysis. Out of possible correlation tests, a two tailed t-test for independent samples was used and two dependent variables were tested in the experiment: Age and Profession.

Age as a dependent variable is said to have some effect on the choice of word order pattern, as was shown in analytical works on clitic placement in Slovak (e.g. Ivanová, to be published). Namely, the older generation shows more acceptability of

the patterns which are viewed as stylistically marked by younger generation (the support for these claims can be found in the work by Ivanová (to be published)).

In Spencer (1973, p. 87), one can find the view that it is possible that the behaviour of producing linguistically relevant intuitions has developed into a specialized skill, no longer directly related to the language behaviour of the speech community. As a consequence, the judgements of linguists may be an artifactual system which reflects the accretion of conceptual organization by linguists. This is why the data from the survey are calculated for the groups of linguists (L) and non-linguists (NL) individually.

In the case of Age as a dependent variable, the following Null hypothesis and Alternative hypothesis were formulated:

Null hypothesis	Alternative hypothesis
There is no difference between the *46 – 75 years old* and the *0 – 45 years old* groups with respect to the dependent variable Value	There is a difference between the *46 – 75 years old* and *0 – 45 years old* groups with respect to the dependent variable Value

In case of Profession as a dependent variable, the following Null hypothesis and Alternative hypothesis were formulated:

Null hypothesis	Alternative hypothesis
There is no difference between the *L* and *NL* groups with respect to the dependent variable Value	There is a difference between the *L* and *NL* groups with respect to the dependent variable Value

4. RESULTS

4.1. Clitic placement in complex clauses

In the given subsection, more attention will be paid to cases when clitic or clitic cluster is licensed not by the predicate in the matrix clause, but by a predicate in a superordinate non-finite clause. In such cases, the clitic which is associated with a verb complex in a subordinate clause can actually be pronounced in a construction with a higher predicate even though it may have no obvious semantic or syntactic connection to that verb (Spencer – Luís 2012, p. 162). Such a phenomenon is referred to as clitic climbing and it is defined as a realization of clitics in a syntactic constituent higher than the licensing predicate (cf. Kulik 2023, p. 211).

The aim in the present subsection is to classify the configurations in which climbing is possible or barred in Slovak. Two types of complex clauses have to be distinguished: if the subject of a matrix verb controls the reference of the PRO subject of its infinitival complement, the verb is called a subject control verb; if the

object of a matrix verb controls the reference of the PRO subject of its infinitival complement, the verb is called an object control verb. The phenomenon of clitic climbing in object control clauses has been widely discussed in many studies on Czech clitics. They have observed that, in the case of infinitive complements, Czech pronominal and reflexive clitics behave in a different way: while clitics can climb out of infinitives which are governed by raising and subject control matrix verbs, some additional restrictions occur in the case of object control matrix verbs (e.g. Dotlačil 2004, Rezac 2005, Hana 2007). On the other hand, there are authors who completely reject possibility of climbing in object control clauses (Junghanns 2002). One such study on clitic climbing proved that some additional restrictions also occur in the case of subject control clauses, (cf. Ivanová (to be published)).

In structures with multiple predicates, clitic climbing can be:

- obligatory: clitic climbing out of infinitival complements of modal verbs is necessary (according to Veselovská 1995, p. 305) and the same applies to complex clauses with phasic matrix verbs like *začať* 'to begin' (Adam 2024, p. 49),
- optional: clitic climbing out of infinitival complements of verbs with subject control is possible and it competes with local placement of clitics within infinitival phrases in these configurations,
- blocked: clitic climbing is blocked in case of some object control clauses (cf. Dotlačil 2004; Rezac 2005; Hana 2007).

The aforementioned rules are not in effect without any exceptions. Even though the local placement of clitics is rather limited with modal verbs, it is possible, for example, in tentative remoteness constructions (cases where a rather vague element of tentativeness, diffidence, extra politeness comes into play):

(3) *To by chcelo pustiť* ***sa*** *do nejakej*
it-NOM COND want-PAST-NEUTR.SG start-INF REFL in some
ucelenej koncepcie.
coherent concept-GEN
'It would be desirable to form some coherent concept.'

(3´) **To by* ***sa*** *chcelo pustiť do nejakej ucelenej koncepcie.*

On the other hand, corpus data bring the evidence on clitic climbing out of object control clauses leading to formation of clitic clusters in which infinitive clitics precede clitics licensed by a matrix verb, as in example (4):

(4) *Pomôž* ***sa*** ***mi*** *obuť.*
help-IMP.SG.2 REFL me-DAT put on the shoes-INF
'Help me put on the shoes.'

One may ask whether the configurations as in example (4) are frequent in Slovak. An initial corpus search for the structure of object control clause with the verb *pomôcť* and climbed reflexive clitic yielded a total of 322 occurrences (5%),

while the search for structure with local placement yielded a total of 5289 occurrences (94%). Corpus data thus show that the given type of structure is relatively infrequent, yet not completely marginal.

Clitic clusters are defined as contact strings of clitics excluding permutation of elements and insertion of non-clitic words. Clitic climbing in subject or object control clauses may give rise to mixed clitic clusters (cf. Kolaković et al. 2022) in which clitics licensed by different matrix VPs occur in adjacent position. However, according to Zimmerling and Kosta (2013), clitic cluster can be formed only by clitics with identical heads which is crucial for distinguishing clitic clusters from occasional word orders like X°— $CL1^X$ | $CL2^Y$ — Y° where two adjacent clitics $CL1^X$ and $CL2^Y$ belong to different syntactic heads X° and Y°.

Slavic languages including Slovak impose grammaticalized constraints on the placement of clitic elements within a clitic cluster. Clitics in clitic clusters are arranged in a rigid order according to language-specific rules called "Clitic Templates" or "Ranking Rules" (Zimmerling – Kosta 2013, p. 179). The internal organization of clitic clusters in Slovak, based on the grammaticalized constraints, can be described as follows.

Table 2. Clitic template of clitic clusters in Slovak

A		B	C						D
Particles		Auxiliary	Pronouns						Connectives
Affirm.	Opt.	Present tense indicative BE-auxiliary	Refl.	Non-argument Dative	Argument Dative	Accusative	Demon.	PPP	Advers.
Že	*By*	*som, si, sme, ste*	*sa, si*	*mi, ti, nám, vám*	*mi, ti, mu, jej, nám, vám, im*	*ma, ťa, ho, ju, nás, vás, ich*	*to, tak, tu, tam*	*s ním, s ňou, k vám*	*však, ale*

Affirm. – Affirmative, Opt. – Optative, Refl. – Reflexive, Demon. – Demonstrative, PPP – prepositional pronoun phrase, Advers. – Adversative

If A°, B° and C° are clusterizing clitics and the fixed order of clitics is [Clitic Phrase A°, B°, C°], no other order like *[Clitic Phrase B°, A°, C°], *[Clitic Phrase C°, A°, B°] should be possible in the canonical position of clusterisation. In accordance with the proposed ordering rules, in constructions with object control verbs reflexive clitics of infinitive Y° can precede the dative clitic of matrix verb X°. This word order pattern X°— $CL1^Y$ | $CL2^X$ — Y° can be occasionally found in the corpus data, not frequently, yet not marginally, e.g.

(5) *Pomohol* ***sa*** ***mu*** *postaviť* *na* *nohy.*
help-PAST-MASC.SG REFL him-DAT stand up-INF on feet-ACC
'I helped him to stand up on his feet.'

The reflexive infinitive clitic *sa* (*obliecť sa*) can even penetrate into the clitic cluster *som jej* licensed by the matrix verb *pomohol*, forming the pattern X°— $CL1^X$ | $CL2^Y$ | $CL3^X$ — Y°:

(6) *Pomohol* ***som*** ***sa*** ***jej*** *obliecť.*
help-PAST-MASC.SG be-PRES.SG.1 REFL her-DAT dress-INF
'I helped her to dress.'

Such linear orderings fully adhere to ordering rules proposed for clitic clusters. On the grounds of the given examples, the following patterns for complex clauses in which both matrix verb and infinitive are cliticized can be sketched:

(i) Non-adjacent placement of clitics conditioned by local placement of infinitive clitic:

(7) *Simona* ***mu*** *pomohla* *vyzliecť* ***sa.***
Simona-NOM him-DAT help- PAST-FEM.SG undress-INF REFL
'Simona helped him to undress.'

(ii) Adjacent placement of clitics which results in mixed clitic clusters X°— $CL1^Y$ $CL2^X$ — Y°; given mixed clusters are either interposed between matrix verb and infinitive (8) or moved in front of matrix verb and infinitive (9):

(8) *Pomohla* ***sa*** ***mu*** *vyzliecť.*
help-PAST-FEM.SG REFL him-DAT undress-INF
'She helped him to undress.'

(9) *Simona* ***sa*** ***mu*** *pomohla* *vyzliecť.*
Simona-NOM REFL him-DAT help- PAST-FEM.SG undress-INF
'Simona helped him to undress.'

(iii) Stacked clitics X°— $CL1^X$ | $CL2^Y$ — Y° which do not form clitic clusters (therefore, ordering rules are not broken in that case); in these cases, the infinitive clitic undergoes partial climbing and its position within the higher clause is disputable – it can be either described as being moved to the third position of the higher clause or as occupying the first position of the infinitive clause as procliticized component:

(10) *Pomôž* ***mi*** ***sa*** *ovládať.*
help-IMP.SG.2 me-DAT REFL control-INF
'Help me to control myself.'

To see whether these constructions are accepted by native speakers, an acceptability judgement survey was conducted. Two types of object control clauses were investigated: with clitic climbing (CC), as in example (11) and with local placement of clitics (LP), as in example (12):

(11) *Kázal* ***sa*** *mi* *vyzliecť.*
order-PAST-MASC.SG REFL me-DAT undress-INF
(12) *Kázal* *mi* *vyzliecť* ***sa.***
order-PAST-MASC.SG me-DAT undress-INF REFL
'He ordered me to undress.'

To prove the frequency of a given structure in the corpus data, a random sample of 100 tokens was drawn from Omnia Slovaca IV Beta and manually annotated. The investigation shows that 70% of annotated examples represent local placement of clitics whereas in 30% of occurrences, the clitics undergo either full (example 13) or partial climbing (example 14) out of infinitive clause:

(13) *Vraj* *si* *mu* ***ich*** *kázala*
apparently be-PRES.SG.2 him-DAT them-ACC order-PAST-FEM.SG
vyhodiť.
throw out-INF
'They say that you ordered him to throw them out.'
(14) *Hneď* *mi* *kázal* ***sa*** *vyzliecť.*
immediately me-DAT order-PAST-MASC.SG REFL undress-INF
'He immediately ordered me to undress.'

The analysis of acceptability ratings brings the following results:

Table 3. Correlations between Age Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC	*0 – 45*: 2.69 *46 – 75*: 2.59 2.64	p = .682	0.08 very small effect	not rejected
With LP	*0 – 45*: 1.37 *46 – 75*: 1.15 1.26	p = .031	0.4 small effect	rejected

Table 4. Correlations between Profession Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC	*L*: 2.48 *NL*: 2.71	p = .404	0.16 very small effect	not rejected
With LP	*L*: 1.18 *NL*: 1.34	p = .127	0.3 small effect	not rejected

Corpus findings can be confronted with survey data on object control clauses. A highly frequent pattern with local placement of clitics is evaluated as highly acceptable by native speakers. On the other hand, the form that is represented only sporadically in the corpus data has a middling rating (2.69 in Group 1 and 2.59 in

Group 2 in the case of depending variable Age, cf. Table 3, and 2.48 in Group 1 and 2.71 in Group 2 in the case of depending variable Profession, cf. Table 4). The results show that high corpus frequency of the pattern correlates with high acceptability rankings, however, rather scarce occurrence of pattern in the corpus data does not automatically lead to low acceptability rankings.

At the same time, the results of statistical analysis show that in the case of pattern with local placement of clitic, the dependent variable Age proves relevant in the case of structure with local placement of clitic component. While the given word order pattern is highly accepted, a sufficient number of respondents from Group 1 consider it acceptable, yet not fully neutral.

There is no statistical difference between Group 1 and Group 2 with respect to Profession as a dependent variable. However, the behaviour of linguists and "naïve" users differs with respect to acceptability span: while the patterns with local placement achieve only the ratings 1 and 2 by linguists, in the non-linguists' group they are rated by full scale from 1 to 5, i.e. the linguists showed significantly greater within-subject consistency than the non-linguists in the given experiment. This proves that fact that linguists may tend to judge strings differently from non-linguists. One possible explanation is that linguists look for reasons behind their acceptance or rejection of a sentence, which takes away spontaneity and makes their judgment processes different from those of naive subjects, who presumably have neither the inclination nor the knowledge necessary to perform this analysis (cf. Schütze 2019, p. 114).

Corpus data on local placement and clitic climbing in object control clauses bring several interesting observations. Clitic climbing can be blocked due to various reasons which are usually described as constraints imposed on clitic components.

Our data confirm the relevance of a constraint labelled as 'Same case, different governors constraint' (cf. Kolaković et al. 2022) which says that clitic climbing might be blocked if two clitics depending on two different matrix predicates have the same case (e.g. in Dative):

(15)	*Kázal*	***mi***	*volať*	***vám***.
	Order-PAST-MASC.SG	me-DAT	call-INF	you-DAT
	'He ordered me to call you.'			
(15a)	**Kázal*	***mi***	***vám***	*volať*.
	Order-PAST-MASC.SG	me-DAT	you-DAT	call-INF

On the other hand, the corpus data do not confirm 'Reflexivity Constraint'. Reflexivity Constraint has been described for cases of multiply embedded infinitive complements. It has been proved that reflexivity of the infinitive that embeds further infinitives plays a crucial role in preventing clitic climbing (cf. Jurkiewicz-Rohrbacher et al. 2017). Apart from stacked infinitives, climbing of reflexive clitics is blocked when the matrix verb has a reflexive counterpart. For example, the

impossibility to move the reflexive clitic into the second position of the matrix verb *nútiť* 'force' is caused by the existence of reflexive verb *nútiť sa* 'force oneself'. The only interpretation of examples like (16) is that the action is understood as being self-oriented (with haplology of reflexive clitics licensed by matrix verb and infinitive):

(16) *Nútila* ***sa*** *zapájať do rozhovoru.*
force-PAST-FEM.SG REFL join-INF into conversation-GEN
'She forced herself to join into conversation.'

To express extroverted meaning, the only solution is to apply local placement of infinitive clitic:

(16a) *Nútila* *zapájať* ***sa*** *do rozhovoru.*
force-PAST-FEM.SG join-INF REFL into conversation-GEN
'She forced (somebody) to join into conversation.'

In structures with the matrix verb *kázať* 'order', reflexive infinitive clitics can climb to a higher clause and occupy the second position. It can be explained on the basis of non-existence of a reflexive counterpart **kázať sa.* Reflexive infinitive clitics can be moved to a higher position as it cannot be confused with reflexive clitics of the matrix verb:

(17) *Na recepcii som* ***sa*** *kázal*
on reception-LOC be-PRES.SG.1 REFL order-PAST-MASC.SG
zobudiť o šiestej večer.
wake up-INF at six in the evening
'I ordered to wake me up at six o´clock at the reception.'

The data also show the relevance of the so-called 'Person Case Constraint' (Bonet 1991), a universal constraint blocking accusative clitics other than the third person when a dative is inserted in the same clitic cluster. First and second person accusative infinitive clitics tend to remain in situ whereas third person accusative clitics can move to a higher position:

(18) *Kto vám kázal tľapkať* ***ma*** *po zadku?*
who you-DAT order-PAST-MASC.SG tap-INF me-ACC on bottom-LOC
'Who told you to pat my bottom?'

(19) *Kázal nám* ***ich*** *prispôsobiť na vašu postavu.*
order-PAST-MASC.SG us-DAT them-ACC adjust-INF on your figure-ACC
'He ordered us to adjust them on your figure.'

4.2 Clitic placement in structures with barriers

In certain syntactic contexts, Slovak clitics appear lower than in clausal 2P. It typically occurs due to informational-structural configurations within the sentence when the topic element occupies the second position, thus preceding the clitics.

(20) *Veterinár bol zvyknutý, že jeho pacienti*
Vet-NOM be-PAST-MASC.SG used to that his patients-NOM
sa *bránia.*
REFL defend-PRES-3PL
'The vet was accustomed to the fact that his patients are defending themselves.'

A similar type of placement may occur after particular sentence constituents which function as a kind of syntactic barrier. Such barriers force clitics to be placed closer to the end of the clause than Wackernagel's Law would lead us to expect.

Following Zimmerling and Kosta (2013), Kosta and Zimmerling (2014), a Barrier can be defined as a syntactic category (a lexical head or a phrase) that has an effect on the position of clitics, namely it can change orientation of a clitic towards the clitic host or move a clitic in a given direction in steps to the right/left of the clitic host. Barrier rules are described as mechanisms that trigger delayed placement of clitics or splitting of clusters.

Several types of barriers are distinguished, namely Obligatory vs. Optional, Grammaticalized (occurring with particular lexical heads) vs. Communicative (phrases with a particular communicative status), Blind (relevant for all types of clitic components) vs. Selective (relevant for certain types of clitic components), Cumulative (when two or more Barriers count as a single Barrier) vs. Undoing (when the second Barrier blocks the effect of the first one).

The NP preceding a clitic, and forming first position within the clause, can be maximally complex (for example, a relative or appositive clause can be added to NP), as long as it still forms one constituent. Given that NP with relative or appositive clauses form the first sentence constituent, a climbed clitic component in these cases follows an intonational break, since it is the position after the first sentence constituent.

(21) *Dvom ďalším prítomným ženám, Faith*
two another present women-DAT Faith-DAT
a Lavender, ***sa*** *podarilo*
and Lavender-DAT REFL succeed-PAST-NEUTR.SG
pozbierať niekoľko zhúžvaných papierových obrúskov.
pick-INF several crumped paper squares-GEN
'Two another women, Faith and Lavender, managed to pick several crumpled paper squares.'

However, NP with relative or appositive clauses can move a clitic one step to the right of the clitic host, functioning thus as a syntactic barrier, e.g.

(22) *Ten,* *ktorý* *sa* *zachránil,*
That-NOM who-NOM REFL save-PAST-MASC.SG
stal ***sa*** *vojakom.*
become-PAST-MASC.SG REFL soldier-INSTR
'The one who saved himself became a soldier.'

To test the possibility of clitic third placement in the object control clauses, and the efficiency of syntactic barrier, the following structures were investigated in the survey:

(i) structures with clitic climbing and clitic third placement:

(23) *Júda,* *vodca* *Makabejcov,* *kázal*
Judah-NOM leader-NOM Maccabees-GEN order-PAST-MASC.SG
sa *modliť* *za* *mŕtvych.*
REFL pray-INF for dead-ACC

(ii) structures with clitic climbing and clitic second placement:

(24) *Júda,* *vodca* *Makabejcov,* ***sa*** *kázal*
Judah-NOM leader-NOM Maccabees-GEN REFL order-PAST-MASC.SG
modliť *za* *mŕtvych.*
pray-INF for dead-ACC

(iii) structures with local placement of clitics:

(25) *Júda,* *vodca* *Makabejcov,* *kázal*
Judah-NOM leader-NOM Maccabees-GEN order-PAST-MASC.SG
modliť ***sa*** *za* *mŕtvych.*
pray-INF REFL for dead-ACC
'Judah, the leader of Maccabees, ordered to pray for the dead.'

Corpus data show that object control constructions with complex initial NP constituents followed by attribute or appositive clause are extremely rare in the corpus. Out of 47 occurrences with verbs *kázať/prikázať*, 63% of examples instantiates local placement of infinitive clitic, 37% brings evidence of clitic climbing (16% of clitics appear in third position, 21% of clitics in second position). Overall, 80% of all clitic component occurs in second position, only 20% of examples exhibits clitic third placement.

The analysis of acceptability ratings brings the following results.

Table 5. Correlations between Age Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC and CL	0 – 45: 2.85 46 – 75: 2.77 2.81	p = .726	0.06 very small effect	not rejected

With CC and no CL	0 – 45: 2.66 46 – 75: 2.33 2.49	p = .232	0.22 small effect	not rejected
Without CC	0 – 45: 1.56 46 – 75: 1.66 1.61	p = .586	0.1 very small effect	not rejected

Table 6. Correlations between Profession Value and Acceptability ratings

Structure	Mean	p-value	Effect size	Null hypothesis
With CC and CL	L: 2.72 NL: 2.85 2.78	p = .574	0.11 very small effect	not rejected
With LP and no B	L: 2.31 NL: 2.65 2.48	p = .251	0.22 small effect	not rejected
Without CC	L: 1.48 NL: 1.68 1.58	p = 0.329	0.19 very small effect	not rejected

Similarly to complex clauses, it is the local placement of clitics that is evaluated as the most acceptable word order pattern. Clitic climbing in object control clauses with complex NP in the first position is evaluated the less acceptable word order variant. At the same time, the lowest score applies to clitic third placement, i.e. the patterns in which infinitive clitic occupies third position after matrix VP.

Statistical analysis also proves no significant difference in acceptability ratings determined by the investigated variables of Age and Profession. Of particular interest are the acceptability ratings by 16 respondents who did not choose any word order pattern as fully acceptable (yet the range of possible patterns was exhaustive and no other patterns could be applied in this cases), the majority of them being non-linguists. It shows that non-linguists often tend to behave more conservatively, are tougher graders (they rated sentences less grammatical overall). At the same time, the ratings of linguists spanned from 1 to 5. Perhaps it shows that linguists are liable to be unconsciously prejudiced by their own theoretical positions, tending to judge in accordance with the predictions of their particular version of grammar (Schütze 2019, p. 113).

5. DISCUSSION

The idea that empirical evidence for theoretical claims should be gathered from multiple sources has become increasingly important for linguistic research of late. The empirical analysis proves that word order variance in clitic placement is larger than expected. Word order variants are both evidenced in the corpus or elicited with relatively high acceptability scores by native speakers. However, the relation between corpus frequency and acceptability ratings is not always straightforward:

1. Higher frequency in the corpus (>70%) entails a high acceptability rating (< 2): this is the case of local placement of infinitive clitics in object control clauses without barriers (70% : 1.26);

2. A high acceptability rating (< 2) does not entail a higher frequency in the corpus (>70%); see e.g. this is the case of local placement of infinitive clitics in object control clauses with barriers (63% : 1.61);

3. Lower frequency (< 30%) in the corpus does not entail a low acceptability rating (>2.5): this is the case of clitic climbing in object control clauses with barriers with no late placement (21% : 2.49);

4. A low acceptability rating (>2.5) entails low frequency (< 30%) in the corpus: this is the case of clitic climbing in object control clauses with barriers and clitics late placement (16% : 2.81).

The investigation also shows that grammaticality is not a dichotomous notion, and grammatical constructions are not simply environments or non-environments for rules; rather they may be environments to a degree and form hierarchies along which different speakers have different acceptability thresholds (cf. Schütze 2019). Grammaticality itself may be thus understood as a gradient phenomenon representing a function of constraint accumulation, i.e. combinations of different grammatical constraints lead to a range of grammaticality levels (cf. Keller 2000, Sorace – Keller 2005, Wasow 2007, etc.).

It is evident that certain positions of clitics seem to be preferred in particular constructions. As a consequence, scholars may consider the less frequent position to be unacceptable. However, the acceptability ratings show that even less frequent patterns are not rejected by native speakers as completely unacceptable. Therefore, the role of the corpus is to determine the circumstances under which rarely occurring clitic positions can be realized in actual usage. The research thus proves that each method adds to better understanding of the studied phenomenon, thus overcoming the possible shortcomings of methods if used independently.

References

ADAM, Nina (2024): *Optimal Clitic Positioning in Czech and Beyond. Clitic syntax, morphology, and prosody from a constraint-based perspective. Dissertation.* Göttingen: Philosophischen Fakultät der Georg-August-Universität Göttingen.

AVGUSTINOVA, Tania – OLIVA, Karel (1995): On the nature of the Wackernagel position in Czech. In: U. Junghanns – G. Zybatow (eds.): *Formale Slavistik.* Frankfurt am Main: Vervuert, pp. 25–47.

BADER, Markus – HÄUSSLER, Jana (2010): Toward a Model of Grammaticality Judgments. In: *Journal of Linguistics*, Vol. 46, No. 2, pp. 273–330.

BERMEL, Neil – KNITTL, Luděk (2012): Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. In: *Corpus Linguistics and Linguistic Theory*, Vol. 8, No. 2, pp. 241–275.

BOOIJ, Geert E. (2012): *The grammar of words: An introduction to linguistic morphology*. 3rd edition. Oxford: Oxford University Press.

BONET, Eulalia (1991): *Morphology after syntax: Pronominal clitics in Romance languages. PhD diss*. Cambridge, MA: MIT.

DOTLAČIL, Jakub (2004): *The syntax of infinitives in Czech*. *Master's thesis*. Tromsø: Universitetet i Tromsø.

FRANKS, Steven – HOLLOWAY KING, Tracey (2000): *A Handbook of Slavic Clitics.* Oxford: Oxford University Press.

HANA, Jirka (2007): *Czech Clitics in Higher Order Grammar*. Columbus: The Ohio State University.

HASPELMATH, Martin (2023): Types of clitics in the world's languages. In: *LTC [Internet]*, Vol. 3, No. 2, pp. 1–59. Available from: https://typologyatcrossroads.unibo.it/ article/view/16057. [cited 2024 Nov. 23]

IONOVA, Anastasiia (2019): *The unbearable lightness of clitics.* Leiden: Universiteit Leiden.

IVANOVÁ, Martina (to be published): Variations on positioning of clitics in Slovak (on the basis of corpus and experimental data) (unpublished manuscript).

JUNGHANNS, Uwe (2002): Clitic Climbing im Tschechischen. In: U. Junghanns (ed.): *Untersuchungen zur Syntax und Informationsstruktur slavischer Deklarativsätze*. Leipzig: Institut für Linguistik der Universität Leipzig, pp. 57–90.

JURKIEWICZ-ROHRBACHER, Edyta – KOLAKOVIĆ, Zrinka – HANSEN, Björn (2017): Web corpora – the best possible solution for tracking rare phenomena in underresourced languages: clitics in Bosnian, Croatian and Serbian. In: P. Bański et al. (eds.): *Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017 including the papers from the Web-as-Corpus (WAC-XI) guest section.* Mannheim: Institut für Deutsche Sprache, pp. 49–55.

KELLER, Frank (2000): *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Ph.D. thesis*. Edinburgh: University of Edinburgh.

KEMPEN, Gerard – HARBUSCH, Karin (2005): The Relationship between Grammaticality Ratings and Corpus Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. In: S. Kepser – M. Reis (eds.): *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Berlin, New York: Mouton de Gruyter, pp. 329–350.

KOLAKOVIĆ, Zrinka – JURKIEWICZ-ROHRBACHER, Edyta – HANSEN, Björn – FILIPOVIĆ ĐURĐEVIĆ, Dušica – FRITZ, Nataša (2022): *Clitics in the wild: Empirical studies on the microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian (Open Slavic Linguistics 7)*. Berlin: Language Science Press.

KOSTA, Peter – ZIMMERLING, Anton (2014): Slavic Clitic Systems in a Typological Perspective. In: L. Schürcks – U. Etxeberria – A. Giannakidou – P. Kosta (eds.): *The Structure of Nominal Expression. Slavic and Beyond*. Berlin: De Gruyter, pp. 441–490.

KULIK, Irenäus (2023): Clitic Climbing without Restructuring in Czech and Polish. In: P. Biskup – M. Börner – O. Mueller-Reichau – I. Shcherbina (eds.): *Advances in formal Slavic linguistics 2021. (Open Slavic Linguistics 8)*. Berlin: Language Science Press, pp. 211–232.

PESCARINI, Diego (2021): *Romance Object Clitics. Microvariation and Linguistic Change*. Oxford: Oxford University Press.

REZAC, Milan (2005): The syntax of clitic climbing in Czech. In: L. Heggie – F. Ordóñez (eds.): *Clitic and affix combinations: Theoretical perspectives*. Amsterdam: John Benjamins, pp. 103–140.

RIEMER, Nick (2009): Grammaticality as evidence and as prediction in a Galilean linguistics. In: *Language Sciences*, Vol. 31, No. 5, pp. 612–633.

ROSEN, Alexandr (2001): *A constraint-based approach to dependency syntax applied to some issues of Czech word order. Ph.D. thesis.* Prague: Charles University.

SCHÜTZE, Carson T. (2019): *The empirical base of linguistics: Grammaticality judgments and linguistic methodology (Classics in Linguistics 2).* Berlin: Language Science Press.

SORACE, Antonella – KELLER, Frank (2005): Gradience in linguistic data. In: *Lingua*, Vol. 115, No. 11, pp. 1497–1524.

SPENCER, Nancy J. (1973): Differences between linguists and nonlinguists in intuitions of grammaticality–acceptability. In: *Journal of Psycholinguistic Research*, Vol. 2, No. 2, pp. 83–98.

SPENCER, Andrew – LUÍS, Ana R. (2012): *Clitics: An Introduction.* Cambridge: Cambridge University Press.

SPROUSE, Jon (2007): Continuous acceptability, categorical grammaticality, and experimental syntax. In: *Biolinguistics*, Vol. 1, pp. 118–129.

SPROUSE, Jon (2008): The differential sensitivity of acceptability judgments to processing effects. In: *Linguistic Inquiry*, Vol. 39, No. 4, pp. 686–694.

SPROUSE, Jon (2009): Revisiting satiation: Evidence for an equalization response strategy. In: *Linguistic Inquiry*, Vol. 40, No. 2, pp. 329–341.

VESELOVSKÁ, Ludmila (1995): *Phrasal Movement and X0-Morphology. Word Order Parallels in Czech and English Nominal and Verbal Projections. PhD thesis*. Olomouc: Univerzita Palackého.

WASOW, Thomas (2007): Gradient data and gradient grammars. In: *Chicago Linguistics Society*, Vol. 43, No. 1, pp. 255–271.

ZIMMERLING, Anton W. (2011): Scrambling types in the Slavic languages. In: *Computational linguistics and intellectual technologies*, Vol. 10, No. 17, pp. 750–766.

ZIMMERLING, Anton W. – KOSTA, Peter (2013): Slavic clitics: a typology. In: *Language Typology and Universals. Sprachtypologie und Universalienforschung*, Vol. 66, No. 2, pp. 178–214.

ZWICKY, Arnold M. – PULLUM, Geoffrey K. (1983): Cliticization vs. Inflection: English N´T. In: *Language*, Vol. 59, No. 3, pp. 502–513.

Sources

Omnia Slovaca IV Maior Beta. Bratislava: Jazykovedný ústav Ľ. Štúra SAV, v. v. i. [cited 2024 Nov. 23].