

Research Article

Martina Ivanová*

Under pressure: Exploring the impact of cognitive factors on clitics placement in L2 Slovak

<https://doi.org/10.1515/opli-2024-0002>

received September 24, 2023; accepted February 28, 2024

Abstract: This article investigates word order of clitics in L2 Slovak and the possible impact of cognitive factors on it. Based on data from the learner corpus, the article makes a first attempt at a more thorough inquiry of ordering patterns in the interlanguage of non-native speakers of Slovak to assess whether cognitive principles of contiguity, relevance, and verb–object bonding possibly have an effect on preferential clitics placement. As proxies for cognitive effects on word order, linguistic and (psycho)linguistic predictors are used, such as proficiency level to consider the possible developmental patterns of word order competence and L1 language group of speakers. The dataset includes 1,051 sentences with the annotated erroneous placement of clitics from learner corpus errkorp-1.0. The data are further annotated manually with respect to relevant variables (type of clitic component, syntactic environment in which this component occurs, expected and actual distance of clitic component from matrix verb in terms of linear dependency segments and syntactic words, actual and expected position of clitic component on right or left periphery of the matrix verb, and proficiency level according to CEFR and L1 language group of speakers). Correspondence analysis based on corpus data demonstrates the relevance of proficiency levels and the irrelevance of the L1 speaking group for preferential orders in the interlanguage. It is thus concluded that the interference hypothesis has little explanatory power for clitic placement in Slovak as L2. It is shown, however, that preferential ordering patterns in the interlanguage can be explained as being regulated by cognitive principles that operate irrespectively of L1.

Keywords: clitics, word order, cognitive principles, interlanguage, Slovak

1 Introduction

The present study examines the word order of clitic components in the interlanguage of students acquiring Slovak as a second language (L2). It tries to provide evidence that preferential word order patterns of clitics are the result of certain cognitive principles rather than the transfer of narrow first language (L1) syntax per se. The research also seeks to confirm or refute Interface Hypothesis according to which “narrow syntactic properties are completely acquirable in a second language, even though they may exhibit significant developmental delays, whereas interface properties involving syntax and another cognitive domain may not be fully acquirable” (Sorace and Filiaci 2006, 340). The word order of clitics in Slovak manifests the phenomenon which illustrates the significance of interface relations. Primarily, clitic placement of Slovak is prosodically conditioned (clitics are prosodically deficient elements so that they do not constitute independent prosodic words and lean on adjacent lexical heads to form prosodic words); however, the cues from the grammar system and information structure can interfere into the linearization patterns of clitic components. In L2

* **Corresponding author: Martina Ivanová**, Institute of Slovak and Media Studies, Faculty of Arts, University of Prešov, Prešov, 08001, Slovakia, e-mail: martina.ivanova@unipo.sk

acquisition, it is exactly interface relations that are persistently problematic in L2 production and often lead to deviations from L1 grammars even at advanced stages of acquisition (Sorace and Serratrice 2009).

In the field of word order acquisition, most studies have focused on various European, especially Germanic and Romance, languages (phenomena like the acquisition of the syntax of negation, e.g. Tomaselli and Schwartz 1990). However, there is a paucity of studies dealing with word-order acquisition in Slavic languages (most studies focus on the relationship between word order and information structure; Cho and Slabakova 2014, Kisselev 2019 for Russian), especially works relying on representative data (from learner corpora or achieved by experiments). The acquisition of word order in Slovak and other Slavic languages is substantially affected by the fact that those languages have great flexibility of word-order patterns depending on the meaning to be conveyed by the sentence structure. Substantial difficulty of word-order acquisition can also be explained on the basis of possible differences existing between word order patterns even in typologically similar Slavic languages.

Since the 1960s, a vast amount of literature in the given field assumes that learners acquire morphosyntactic structures in a certain predictable, sequential way. Learners' linguistic competence develops and progresses along 'natural order' (e.g. Ellis 2008, Ortega 2009). To identify those development patterns, certain intralingual (as the influence of L1) and extra-lingual (e.g. the age of learners, learning environment, etc.) factors have been taken into consideration. The acquisition of different word order phenomena within L2 (including various ordering templates) has been explained through theoretical positions of Universal Grammar (predictability of word order is explained by some underlying principles of Universal Grammar), e.g. Clahsen and Muysken (1989), Hulk (1991); Processability Theory (which states that the teachability of L2 structures at certain stage of interlanguage development is constrained by the same processing restrictions that determine the developmental sequences of natural language acquisition) e.g. Pienemann (2005); Competition Model (suggesting that L1 appropriate processing strategies and cue hierarchies can carry over into L2 processing), e.g. Bates and MacWhinney (1982, 1989); or cognitive concepts (which work with conceptual word-order principles underlying linearization patterns in particular language), e.g. Jiang (2009).

We proceed as follows. In Section 2, cognitive factors possibly underlying the word order acquisition are analysed. Section 3 provides the definition of clitic components and their classification in Slovak. In Section 4, our sampling method and the database which builds on data from the acquisition corpus of Slovak as L2 are presented. Section 5 is the main part of the article and is devoted to the errors in clitics placement with respect to the types of clitic components, proficiency levels, clitic systems of L1, and syntactic environments. Section 6 summarizes the results and presents conclusions.

2 Cognitive principles underlying the word order acquisition

At the very beginning of the investigation of word order acquisition, a simple question can be asked: Why do the learners prefer certain word order structures over others, often irrespectively or counter-intuitively with regards to the dominant or preferred word order structures existing in their native languages? Are there any word order templates frequently employed in the interlanguage and if yes, how can their existence be explained? Are there any cognitive principles underlying constituent orders which either motivate or constrain word order preferences? Is it possible to reveal the cognitive mechanisms that govern the preferential choice of specific constituent orders by L2 learners?

If there is a positive answer to the aforementioned questions, then it can be argued that those constituent orders which mirror certain cognitive principles occur with greater frequency in the interlanguage of L2 learners.

2.1 The principle of contiguity

The cognitive principle called 'contiguity' represents the concept according to which ideas, memories, and experiences are linked when one is frequently experienced with the other. There are two different types of

contiguity: semantic and formal (positional). Semantic contiguity can be described as the closeness of two concepts which enables the conceptualizer to evoke one concept by the other. Formally, contiguity is realized as syntagmatic or morphological bondedness. The concept of bondedness is described by Lehmann as “the intimacy with which it [a sign] is connected with another sign to which it bears a syntagmatic relation,” and is understood as an indication of the degree of grammaticalization. “The degree of bondedness of a sign varies from juxtaposition to merger, in proportion to its degree of grammaticality” (Lehmann 2002, 131). In language theories, contiguity is described as the basis of syntagmatic relations (which are formed by contiguity) which oppose associate relations (which are formed by similarity) in Saussure’s synchronic two axes theory (Koch 1999, 142).

The cognitive preference for contiguity can be explained as being grounded in prominent efficiency-based accounts of word order which include various locality principles. Two locality principles in word order typology have been established in the theoretical studies. The first principle is dependency locality (DL), according to which languages tend to order words to reduce the overall distance between syntactically related words (Hawkins 1994). DL can be justified in terms of parsing efficiency, memory efficiency, and general communicative efficiency (Hawkins 1994; Gibson 1998; Hahn et al., 2020). The second locality principle is information locality (IL), whereby languages favour placing elements together when they are syntactically related or contextually informative about each other (Qian and Jaeger 2012).

Cliticization of formal elements representing certain grammemes or arguments of their matrix verbs and their discontinuous placement into the second position, distant from their matrix verb, can be considered as a violation of the DL principle as it includes the subordination of the clitic item under its prosodic host and (possible) disintegration from its syntactic or morphological governor. On the other hand, the verb-adjacent placement of clitic component minimizes total dependency length in accordance with the Dependency-Length Minimization hypothesis.

Research question 1: Do learners prefer those constituent orders which maintain adjacency of syntactically or morphologically bonded elements to enable short dependencies of the matrix verb and its clitic component in usage?

2.2 The principle of relevance

Speaking of the relevance principle, not only the proximity of language elements but also their constituent order should be taken into consideration. The principle of relevance can be ascribed to clitics components functioning within the inflection domain (verbal auxiliary clitics marking person and number in past tense and reflexive clitics marking voice in Slovak) and derivation domain (reflexive clitics functioning as derivational means for derivation of reflexive verbs).

With respect to grammatical markers, research by Bybee et al. (1990) scrutinized whether there is (i) a general preference for postposing grammatical material or (ii) a greater tendency to fuse elements that are already postposed. Their survey of a set of grammatical markers in a 71-language sample showed both.

Grammatical markers (Bybee et al. 1990, 5):

	All	Free	Bound
Preposed	34%	48%	52%
Postposed	66%	20%	80%

The research shows that postposed grammatical markers outnumbered preposed markers two to one (66%:34%), which explains the strong attraction of verbal auxiliary clitics as well as reflexive voice clitics to the right periphery of the matrix verb. The same tendency can be ascribed to reflexive clitics serving as valence markers (changing the number of arguments and valency of the reflexive verb derived from its non-reflexive counterpart) in Slovak (e.g. anticausative, deaccusative verbs). Their attraction to the position right to the

matrix verb mirrors Universal Affix Ordering (Bybee 1985) according to which valence affixes succeed verb stem.

At the same time, both verbal auxiliary clitics and reflexive clitics in Slavic languages display grammaticalization processes from word to clitic and, in tendency, to postfix with the position on the right periphery after other derivational and inflectional suffixes. In a prototypical grammaticalization cline, according to Hopper and Traugott (2003, 7), clitic is understood as an intermediate stage between grammatical word and inflectional affix: content item > grammatical word > clitic > inflectional affix. However, this idea has been challenged by Haspelmath (2023), arguing that most types of clitics do not become affixes, and most types of affixes did not pass through a clitic stage. Especially ‘Wackernagel clitics’ (as found in Slovak) are unlikely to become affixes, as affixation requires a stable host whereas Wackernagel clitics are typical of promiscuity, i.e. word class selectivity. Postfixal status of certain types of clitics is evidenced in several Slavic languages (e.g. auxiliary postfix in past tense is evidenced in Polish, reflexive postfix in Russian). It confirms the tendency anticipated in research by Bybee et al. (1990) for postponed elements to be bonded. According to the principle of relevance (Bybee 1985), morphemes of greater relevance to the verb will generally occur closer to the verb stem; those of lesser relevance will generally occur at a greater distance. Reflexive clitics do not modify the verb semantics in a substantial way; the only change they bring into construction is to signal that the terminal point of the causal chain meaning is ascribed to the subject referent (in canonically reflexive verbs) or to trigger the implicature that the use of its lexical causative counterpart is too strong (in anticausative verbs).

These tendencies can be explained cognitively: the right peripheral position of inflectional formatives and derivation formatives with lesser relevance for the semantics of the verbal unit facilitates their processing through the effects of psychological primacy and recency and better outward indexicality towards the other parts of the sentence (Dressler et al. 1987, 7, cited in Haspelmath 1993, 292).

Research question 2: Do learners prefer those constituent orders in which clitic elements from inflection or derivation domains are placed on the right periphery of their matrix verbs?

2.3 The principle of verb–object bonding

According to the principle of verb–object bonding (Tomlin 1986), the relation between the verb and the object is stronger than that between the subject and the verb. The distribution of constituent orders across language families shows certain preferences with SOV being dominant, with SVO a close second, and VSO a distant third. The remaining orders (OSV, VOS, and OVS) are extremely rare (Dryer 2013). Furthermore, languages are known to change from SOV toward SVO, but the reverse change is much rarer (Gell-Mann and Ruhlen 2011).

There were several explanations offered for this fact, namely that SOV may be well-suited for describing non-reversible events, but it might be less well-suited for describing reversible events (Hall et al. 2013). Other explanations generally assume that what makes SOV problematic is the potential for confusability when two human nominals both precede the verb, which leads to role conflict (*ibid.*).

An alternative explanation for this issue is offered by the theory of uniform information density (UID). The UID hypothesis suggests, “language producers unconsciously endeavour to keep the rate of information transmission as close to constant as possible when speaking” (Maurits et al. 2010, 1587). In the light of the presented cognitive and pragmatic tendency, closeness to this UID ideal can explain the preferred distribution of certain word orders. According to this theory, object-first word orders deviate from uniformity because the first word (the object) carries a disproportionate amount of information as objects are predictive of very few subjects or verbs. For instance, the object word ‘water’ implies only a few possibilities for verbs (e.g. ‘drink’), which in turn restricts the subjects (e.g. to living things). By contrast, hearing the verb ‘drink’ implies many possibilities for objects (e.g. ‘water’, ‘coffee’, ‘cola’, ‘juice’, etc.; Maurits et al. 2010, 1592). This may explain verb–object order preference over object–verb order.

Research question 3: Do learners prefer constituent orders in which subjects precede verb and objects follow the verb?

3 Clitics in in Slovak

3.1 Definition of clitics

Among factors determining the linearization of the components in the sentence structure in Slovak prosody, grammar and functional sentence perspective have been mentioned (e.g. Mistrik 1966). Informational structure, incarnated in the principles of functional sentence perspective, is considered the major principle of Slovak word order, whereas prosody and grammar play the role of accompanying factors as they are only employed when certain types of units appear in the sentence (namely clitic components with respect to prosody and attributives with respect to grammar).

Prosody relates to the word order realization of phonologically non-independent elements devoid of word stress called clitics which cannot be the sole elements of an utterance and cannot be realized freely, i.e., in different sentence positions depending on pragmatic and discourse function, but their position within sentence structure is determined phonologically.

However, phonological criteria like prosodic dependency or deficiency are not the most decisive features of these elements. As Haspelmath (2022, 20) states, clitics and affixes are very similar in that they are prosodically deficient, and there is no simple phonological criterion distinguishing affixes from words (including both free words and clitics). In many theoretical studies, two relevant characteristics of clitics are emphasized: bondedness, i.e. non-occurrence in isolation (which sets them apart from free words), and non-selectivity, i.e. lack of word-class selectivity or promiscuity (which differentiates them from affixes; Haspelmath 2022).

Slovak belongs to those languages which follow Wackernagel's Law and its clitic elements belong to the category of second-position clitics (CL2). There are two features of 2P syntax: (a) that a combination of two (or more) phrasal categories X, Y preceding clitics in a CL2 language should be ungrammatical; (b) that the XP-position in any clause type with CL2 is not reserved for any particular syntactic category (e.g. noun phrase) and does not express any particular grammatical relation (e.g. subject; Zimmerling 2015).

3.2 Classification of clitics

In many theoretical works, unstressable and unstressed clitic components should be differentiated, the former labelled as clitics tantum and the latter as volatile clitics.

Clitics tantum can be defined as prosodically deficient unstressable elements that are always unstressed, independently of the context in which they are realized, thus being unable to be focused and moved into the initial position. Contrarily, volatile clitics can be defined as prosodically unstressed elements that can have phonological autonomy under certain contextual conditions. Clitics tantum which can only occur in the post-initial position is often referred to as 'pure sentential clitics' (Avgustinova and Oliva 1995) or 'constant clitics' (Rosen 2001, Hana 2007), whereas volatile clitics which may as well as not occupy post-initial position are labelled as semi-clitics (Avgustinova and Oliva 1995). A similar difference can be found in the work by Junghanns (2002) who differentiates lexical clitics for which clitic status represents an inherent part of their lexical 'equipment', and phonological clitics for which their clitic status is formed 'in the phonological part of the grammar usage' (elements that optionally undergo phonological cliticization). Functionally, clitics can be classified into several subtypes, depending on the grammatical and lexical status of the clitic component.

3.2.1 Auxiliary clitics

Auxiliary clitics in Slovak are person and number markers in one-participle forms in the past tense and the pluperfect tense. L-participle, having semantically developed from a present perfect construction, is the only past tense construction in modern contemporary Slovak; pluperfect tense forms are stylistically and

temporally marked forms. Slovak maintains the full form of the auxiliary as a sentential clitic in the first and the second person; in the third person, the auxiliary is omitted. As opposed to Czech, grammaticalization in the second person and formation of reduced forms with *-s* is not evidenced in Slovak.

- (1) Tos mi tak ale neřekl.
 That-ACC + be-AUX-2SG me-DAT so but NEG-tell-PST-MASC.SG
 ‘You didn’t tell me that’
 (Clancy 2010, 166)

- (1') To si mi tak ale nepovedal.
 That-ACC be-AUX-2SG me-DAT so but NEG-tell-PST-MASC.SG
 ‘You didn’t tell me that’

Occasionally, we can find rare examples of fusion in particular Slovak dialects close to Czech, e.g. *tys došeu* (Malacky), *tys = ty si prišiel* ‘you have come’.

3.2.2 Argument clitics

Argument clitics are represented by weak or short pronominal forms coding both direct and indirect objects or adjuncts. Among argument clitics, the forms of non-prepositional personal (*ma* ‘me-ACC’, *ta* ‘you-SG-ACC’, *ho* ‘him-ACC’, *ju* ‘her-ACC’, *nás* ‘us-ACC’, *vás* ‘you-PL-ACC’, *ich* ‘them-ACC’, *mi* ‘me-DAT’, *ti* ‘you-SG-DAT’, *mu* ‘him-DAT’, *jej* ‘her-DAT’, *nám* ‘us-DAT’, *vám* ‘you-PL-DAT’, and *im* ‘them-DAT’) and demonstrative pronouns (*to* ‘it-NOM/ACC’, *tu* ‘here’, *tam* ‘there’, and *tak* ‘so’) can be mentioned. Within the group of personal pronouns, two groups can be differentiated, namely clitic short forms of personal pronouns which stand in opposition to long accented forms (*ma* – *mňa* ‘me-ACC’, *ta* – *teba* ‘you-SG-ACC’, *ho* – *jeho* ‘him-ACC’, *mi* – *mne* ‘me-DAT’, *ti* – *tebe* ‘you-SG-DAT’, and *mu* – *jemu* ‘him-DAT’) and unparalleled forms which can be used as either clitics or as accented full forms (*ju* ‘her-ACC’, *nás* ‘us-ACC’, *vás* ‘you-PL-ACC’, *ich* ‘them-ACC’, *jej* ‘her-DAT’, *nám* ‘us-DAT’, *vám* ‘you-SG-DAT’, and *im* ‘them-DAT’).

The clitic status of unparalleled forms can be proved by their realization within the clitic cluster. On the basis of rule (b) from Zimmerling (2008), element X between 1P and clitic component is a clitic:

- (2) Oni nám ho vzali. (Omnia Slovaca IV Maior, OSIVM)
 they us-DAT it-ACC take-PST-PL
 ‘They took it from us’

As the unparalleled form *nám* is interposed between 1P (*Oni*) and the permanent clitic (*ho*), it can be considered a clitic.

Demonstrative monosyllabic pronouns behave as volatile enclitics; i.e. they can occupy the initial position in the stressed form, or under certain conditions, they become prosodically dependent and behave as clitic elements. First, there is an additional pronominal form, the neuter demonstrative *to* which can replace the expected *ho* and appears in expletive contexts (Franks and Holloway King 2000, 101). Second, there are also other demonstrative pronoun forms with a range of adverbial functions like *tam*, *tu*, and *tak* which have strong counterparts *tamto*, *tuto*, and *takto* in emphatic contexts.

- (3) Zdá sa ti to rozumné? (OSIVM)
 seem-PRS-3SG REFL you-DAT it-NOM reasonable
 ‘Does it seem reasonable to you?’

Apart from bare pronominal forms, clitic nature can be ascribed to monosyllabic prepositional forms of personal and demonstrative pronouns (*k nám*, *k vám*, *k nej*, *s ním*, *s ňou*, and *s tým*), which behave as volatile clitics depending on the context.

- (4) Ja som s ním však nikdy žiaden problém
 I 1SG with him-INSTR however never any problem-ACC
 nemal. (OSIVM)
 NEG-have-PST-MASC.SG
 ‘However, I have never had any problem with him’

3.2.3 Reflexive clitics

Reflexive clitics in Slovak are polyfunctional elements that occur as argument reflexives within the syntactic domain, lexical reflexives within the derivational domain, or grammatical reflexives within the morphological domain. The exact syntactic status of the short pronouns has been frequently discussed in theoretical studies just as the boundary between syntactic and lexical reflexives. Slovak argument reflexive clitics *sa*, *si* represent short pronominal forms which make a complementary distribution with the corresponding long forms *seba*, *sebe*. However, this is not the case for lexical and grammatical reflexives. Independent of their status, all reflexive clitics show similar word order behaviour. Reflexive *sa* always behaves as an independent clitic as opposed to Russian which attests the synthetic passive reflexive constructions.

- (5) Vorota odkryvalis' storožem
 gate-NOM.SG open-PST-3PL-REFL watchman-INSTR.SG
 ‘A/The gate was opened by a/the watchman’
 (Gehrke 2018, 5)
- (5') Brána sa otvárala (strážcom).
 gate-NOM.SG REFL open-PST-FEM.SG (watchman-INSTR.SG)
 ‘A/The gate was opened’

3.2.4 Be-forms clitics

The clitic status of *byť* ‘be’ is disputable and there is a disagreement as to the clitic nature of copular *be* in Slavic languages. According to *Nový encyklopedický slovník češtiny*, only auxiliary *be* is assigned a clitic status (Uhlířová et al. 2017). The same opinion can be found by Franks and Holloway King (2000). On the other hand, in the work by Avgustinova and Oliva (1995), only auxiliary *be* in the present tense within the passive construction is described as a clitic element. According to Junghanns (2002), the forms of the verb *be* can behave as phonological clitics, yet it is not specified if this assumption holds only for auxiliary *be* or also for copular (and existential) *be*. According to Kosek (2011), only auxiliary forms of past conditional and antepreterite (*byl/a* for Czech) and present forms of non-auxiliary *být* can acquire clitic status.

In the present approach, the status of unstable clitic component is assigned to monosyllabic forms of the verb *byť* independently of its lexical status (copula, lexical verb), based on the postulated rule [1P – X – C]: element X (*je*) between 1P (*skoro*) and clitic component (*mi*, *ťa*) is a clitic (*je* = clitic; Zimmerling 2008):

- (6) Skoro je mi ťa ľúto. (OSIVM)
 almost be-PRS-3SG me-DAT you-ACC sorry
 ‘I feel almost sorry for you’
- (6') *Skoro bolo mi ťa ľúto
 almost be-PST-NEUTR.SG me-DAT you-ACC sorry
- (6'') Skoro mi ťa bolo ľúto
 almost me-DAT you-ACC be-PST- NEUTR.SG sorry
 ‘I felt almost sorry for you’

3.2.5 Particle clitics

Among particle clitics, elements expressing the relation to the content of the entire clause can be counted. Similarly to Czech, which is described in the work by Spencer and Luis (2012), Slovak has particle clitics which bear a range of discourse functions.

Conditional single particle *by* is used to form present and past conditional forms. In the present form, it is combined with one-participle of the main verb, and in the past tense, *by* combines with one-participle of the main verb and additional one-participle form of the verb *byť*. This additional one-participle form behaves as a semi-clitic, and it can occur in the second position along with other clitic forms or can behave as a prosodic host in the first position.

(7) Určite by som pokračoval (OSIVM)
certainly COND be-AUX-1SG continue-PST-MASC.SG
'I would certainly continue'

(8) Určite by som bol pokračoval.
certainly COND be-AUX-1SG be-AUX-PST-MASC.SG continue-PAST-MASC.SG

(8') Bol by som určite pokračoval
be-AUX-PST-MASC.SG COND be-AUX-1SG certainly continue-PST-MASC.SG
'I would have certainly continued'

3.2.6 Clitic clusters

Clitic clusters can be defined as contiguous strings of two or more clitics taking an adjacent position, such that (a) the clitics are arranged in a rigid order and are not permutable, (b) insertion of non-clitic words is not allowed, (c) all clitics have one and the same prosodic host they are attached to; (d) all clitics belong to one and the same projection and not to two adjacent phrases; (e) all clitics can move as a unit; (f) adjacent clitics in a group do not form new lexical items (Zimmerling 2008, 20).

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 and Kosta 2013, 179). The internal organization of clitic cluster in Slovak, based on the grammaticalized constraints, is described in Table 1.

Table 1: Clitic template of clitic clusters in Slovak

A		B	C						D
Particles		Auxiliary	Pronouns						Connectives
Affirm	Opt	Present tense indic. BE-auxiliary	Refl	Non-argument dative	Argument dative	Accusative	Demon.	PP	Advers
že	by	som, si, sme, ste	sa, si	mí, ti, nám, vám	mí, 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

4 Methodological aspects of research

The language material used in the current study comes from the learner corpus of Slovak language called errkorp-1.0. In its current state, the corpus comprises 347,393 tokens and 47,439 words. It is a collection of written texts from non-native speakers of Slovak as L2 with different L1 (37 languages) of all language

proficiency levels defined by CEFR (A1–C2 levels). The corpus contains metadata on both authors and texts. The corpus contains complete metadata related to the learner (age, sex, native language/s, languages spoken by a learner, etc.), the topic of the text, or the circumstances where the text was produced. The corpus has internal annotation, namely error annotation as the explicit and transparent way of marking errors in texts. It contains 49 tags, and it is structured in two levels of information: (i) area or level affected (one segment, morpheme, word, word combination, and text); (ii) error category (and subcategories in some cases).

Errors concerning non-adequate clitic placement are tagged as ‘order’. All sentences with erroneously used enclitic components were transcribed into Excel and were assigned annotation tags reflecting the investigated variables. The following parameters were taken into consideration during annotation.

Type of clitic component:

Reflexive clitics: R = reflexive component

Auxiliary clitics: GM = auxiliary *byť* (grammatical morphemes *som*, *si*, *sme*, and *ste* coding person and number grammemes in past tense forms)

Particle clitics: K = conditional morpheme *by*

Argument clitics: P = short form of personal pronoun, D = monosyllabic demonstrative pronoun

Be-forms: C = non-auxiliary forms of *byť* used either as parts of passive participle or as copulas

Clusterized enclitics: CC = clitic cluster

Syntactic environment:

JV = clitic component occurring in a simple sentence or in a second clause within a compound sentence not introduced by a conjunction

PS = clitic component occurring in a second clause within a compound sentence introduced by a conjunction

VHH = clitic component occurring in a postponed main clause of a complex sentence

VVH = clitic component occurring in an initial subordinate clause of a complex sentence

HVV = clitic component occurring in a subordinate clause of a complex sentence

HVH = clitic component occurring in a main clause after interposed subordinate clause within a complex sentence

IC = clitic component within subject or object control infinitive construction

Expected proximity to lexical/syntactic host with respect to number of linear dependency segments

(LDS): 0 = zero LDSs, 1 = 1 LDS, 2 = 2 LDSs...

Expected proximity to lexical/syntactic host with respect to number of syntactic words: 0 = zero words, 1 = 1 words, 2 = 2 words...

Actual proximity to lexical/syntactic host with respect to number of linear dependency segments

(LDS): 0 = zero LDSs, 1 = 1 LDS, 2 = 2 LDSs...

Actual proximity to lexical/syntactic host with respect to number of syntactic words: 0 = zero words, 1 = 1 word, 2 = 2 words...

Expected position in relation to lexical/syntactic host: PreV = preverbal position, PostV = postverbal position

Actual position in relation to lexical/syntactic host: PreV = preverbal position, PostV = postverbal position

The entire database is published online by Ivanová (2023) and is freely accessible. Overall, 1,051 sentences with clitic components annotated as erroneously used were analysed. Error identification is always based on some kind of reconstructed learner utterance (target hypothesis, TH). In the present research, data on the learner's language proficiency level (including six categories: A1, A2, B1, B2, C1, and C2) and their first language (L1) are utilized to categorise learners into Slavic and non-Slavic groups.

The current study aims to analyse word order development concerning clitic placement in the interlanguage of learners of Slovak as L2 focusing on three aspects:

- (i) ability to place the clitic component to 2P position even though the placement violates the adjacency of clitic component and its matrix verb (principle of contiguity);

- (ii) ability to put the clitic component into preverbal position despite canonical word order of auxiliary and reflexive clitic components on the right periphery of the matrix verbs (principle of relevance) and despite the canonical SVO word order (principle of verb–object bonding);
- (iii) the rivalry between topic elements and clitic components to occupy 2P position.

5 Results

To quantify the most frequent errors of word order in the written productions of students learning Slovak as L2, the method of correspondence analysis (CA) was used. CA can be defined as an exploratory technique that reveals frequency-based associations in complex corpus data. The technique visualizes these associations to facilitate their identification in the form of a Euclidean cloud (resembling ‘rice thrown onto a board or the holes made by darts on a dartboard’) which depicts degrees of correlation and variation through the relative proximity of data points. At the same time, CA uses the weighted averages of the profiles to compensate the bias caused by infrequent features which would have a disproportionate effect if all were taken equally (Glynn 2014, 446).

The method makes it possible to map the association between the investigated values based on the relative frequency of word order errors in the texts (e.g. type of clitic component and type of syntactic environments in which error occurs, the attraction of clitic placement to pre- or postverbal position with respect to individual proficiency levels, etc.). Values in the graphs visualizing CA data which are close to each other show mutual attraction while values that are far apart on a plot are repulsed from each other. However, CA requires a reasonable hypothesis about where to look; it is “a tool for digging in the data to look for patterns and correlations, but it certainly helps if one knows where to dig” (Glynn 2014, 445).

5.1 Errors with respect to type of clitic component and proficiency level

The first value to be investigated is the type of clitic component and the distribution of certain clitic types with respect to proficiency levels. Figure 1 displays the relative positions of both the row vectors – types of clitic components – and the column vectors – the proficiency levels.

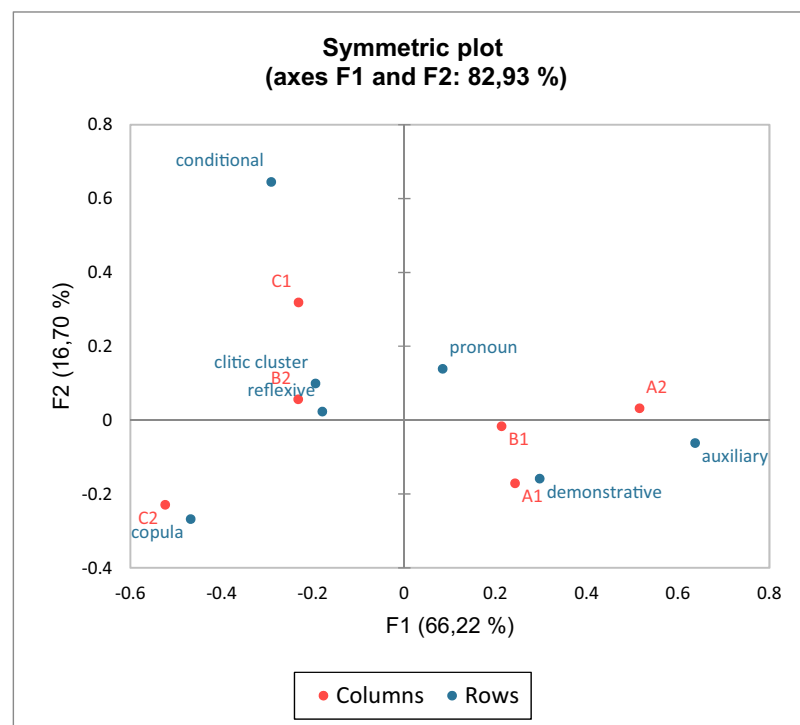


Figure 1: Overall distribution of errors with respect to proficiency levels.

We can summarize the interpretation of the plot as follows:

- conditional clitic is distinctively associated with C1,
- clitic cluster and reflexive clitics are distinctively associated with B2,
- copula is grouped with C2 (higher error ratio of non-auxiliary *be*-forms at C2 level is explained in section 6.4),
- demonstrative is grouped with A1, and
- B1 is associated with auxiliary.

To explain the attractions revealed by CA, the relative frequencies of clitic components in the texts from individual proficiency levels presented in Table 2 are compared with error rate at these levels.

Table 2: Relative frequency of selected clitics in errkomp-1.0 according to proficiency levels

spk.level_ERR	AUX Rel (%)	COND Rel (%)	DEM Rel (%)
A1.1	13,911,62	2,104,28	5,143,79
A1.2	36,908,65	1,101,14	4,119,09
A2.1	38,924,44	3,275,44	6,665,14
A2.2	51,930,70	6,033,44	8,963,97
B1	57,002,26	5,985,80	9,567,60
B2	31,600,50	5,949,32	14,426,68
C1	45,649,67	4,268,16	12,583,71
C2	31,512,56	1,979,61	7,665,22

The results suggest several meaningful discrepancies in lower versus higher proficiency levels concerning the ratio of particular types of clitics and their erroneous usage.

First, lower-level learners display an overall underuse of demonstrative clitics; however, erroneous usage of demonstrative clitics reaches the highs at A1. The possible explanation for this tendency lies in the fact that demonstrative items can be realized either as accented forms with corresponding freedom of occurrence, limited only by functional sentence perspective, or as unaccented clitics confined to second position. Two tendencies for erroneous placement of demonstratives are observable in the data.

- (i) Speakers often use non-focused argument demonstratives in postverbal position which can be explained by the principle of verb-object bonding:
- (9) Určite nevnímam to ako pozitívum. (ru, C1)
 certainly NEG-perceive-PRS-1SG it-ACC as positive-ACC
 ‘I certainly do not perceive it as a positive fact’
 TH: Určite to nevnímam ako pozitívum
- (ii) Speakers often place anaphoric demonstratives into initial position even in non-emphatic contexts which can be explained by the unstable status of demonstratives varying between full form and clitic and low pragmatic competence of speakers:
- (10) (Moja krajina Afganistan sa nachádza v srdci Ázii)
 Tam žijú Afganci (ps, B2)
 there live-PRS-3PL Afghans-NOM
 ‘(My country Afghanistan lies in the heart of Asia.) The Afghans live there’
 TH: Žijú tam Afganci

Second, learners use an increased amount of conditional clitics in intermediate levels, whereas erroneous usage of conditional marker *by* is overrepresented at advanced proficiency level (C1). This mismatch can be explained by two general factors:

- (1) higher error ratio of certain types of clitics can be linked to growing syntactic complexity of texts at higher proficiency levels which is reflected in more complex clause structures;
- (2) higher error ratio of certain types of clitics mirrors growing demands on information management reflected in word order preferences at prosodic-pragmatic interface.

The data show that erroneous usage of conditional clitics often occurs in sentence structures with multiple themes or with sentence adverbials functioning as disjuncts which strongly incline to be placed in the thematic zone of utterance (e.g. adverbials referring to source of knowledge) and may cause delayed placement of clitic component in the texts as communicative barriers:

- (11) Oblečenie značky Medicine podľa reklamy
 clothing-NOM mark-GEN Medicine according to advertisement-GEN
 by malo vyliečiť závislosť od
 COND should-PST- NEUT. SG heal-INF dependency-ACC on
 smartfónu. (pI, C1)
 smartphone-GEN
 ‘The clothing with Medicine mark should heal, according to the advertisement, the dependency on smartphone’
 TH: Oblečenie značky Medicine by podľa reklamy malo vyliečiť závislosť od smartfónu

These discrepancies between relative frequencies of clitic components in the texts and ratio of their erroneous usage in particular proficiency levels can be labelled as ‘frequency/error rate’ mismatch.

On the other hand, frequency distribution of past tense auxiliary clitics and their error ratio in particular proficiency levels show correspondence. Higher error ratio of auxiliary clitics at B1 level can be explained by overall distribution of past tense structures at this proficiency level.

5.2 Errors with respect to types of syntactic environments and proficiency levels

The second value to be investigated is the possible attraction between the erroneous usage of clitics components in certain syntactic environments with respect to proficiency levels. Figure 2 shows the relative positions of both the row vectors – types of syntactic environments – and the column vectors – the proficiency levels.

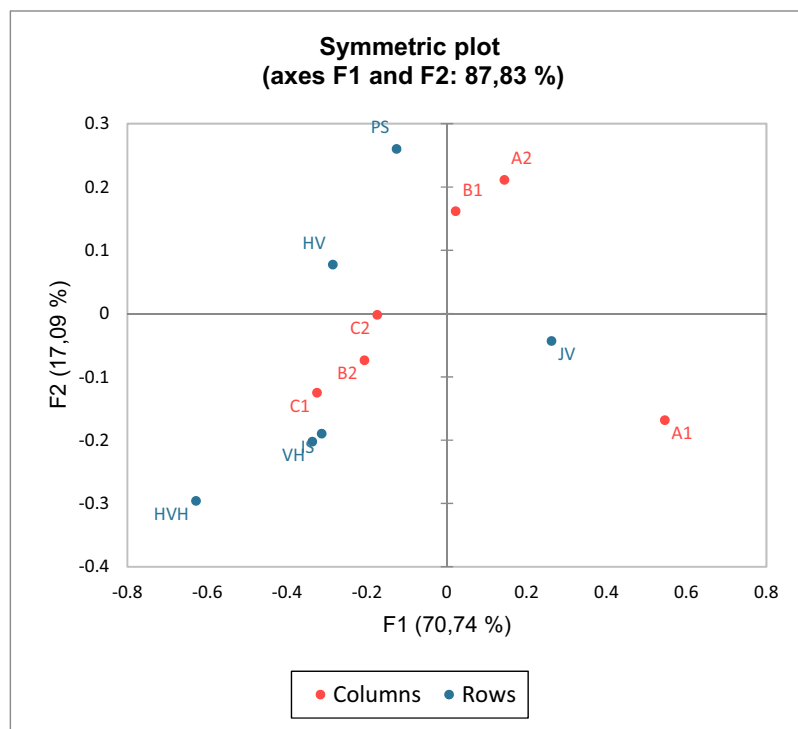


Figure 2: Overall distribution of types of syntactic environments with respect to proficiency levels.

We can summarize the interpretation of the plot as follows:

- syntactic environments described as HVH, VH, and IS are grouped with C1 and B2;
- syntactic environment described as JV is distinctively associated with A1.

The results point to an important fact concerning the level of syntactic complexity of texts at different proficiency levels. Erroneous usage of clitics in simple clauses is typical of low-level learners (A1) which may be caused by higher ratio of simple clauses in these levels, as the investigation for Slovak proved (Ivanová et al. 2021).

Growing syntactic competency leads to adhering more rigorously to word order rules in simple clauses at higher proficiency levels. On the other hand, more complex syntactic structures used at higher proficiency level attract more frequent difficulties with clitics placement. This tendency can be explained on the basis of the theory labelled as Barrier Theory which seeks to explain deviations from the basic order (2P in case of Slovak) as the result of different ‘barriers’ that hinder proper clitic placement. Barrier can be defined as certain type of sentence-initial groups or lexical heads hosting the clitic which have the properties (grammatical or communicative) due to which clitic element is moved from 2P to the right of this position (Zimmerling and Kosta 2013).

Grammatical barriers occur in certain types of syntactic environments (e.g. in main sentence with an interposed subordinate clause). In canonical word order templates, Slovak clitics stick to 2P position even when preceded by syntactic units which necessarily induce an Intonational Phrase boundary (Nespor and Vogel 1986). In texts produced by native speakers, clitics show default placement immediately after the IP boundary:

- (12) Dieťa ktoré držala žena v náručí, sa
 child-NOM which hold-PST-FEM.SG woman-NOM in arms-LOC REFL
 odrazu prebralo (OSIVM)
 suddenly wake up-PST-NEUT.SG
 ‘The child that woman was holding in her arms woke up’

In such cases, clitics are prosodically parsed as proclitic on what follows them.

However, the data show that non-native speakers often employ delayed placement in these types of syntactic structures so that the clitic component is not placed immediately after the I-phrase boundary but after the second syntactic phrase:

- (13) Knihy ktoré má zaujímajú musela
 books-NOM which me-ACC interest-PRS-3PL have_to-PST-FEM.SG
 Som vymeniť za učebnice. (sr, B2)
 be-AUX-1SG exchange-INF for textbooks-ACC
 ‘I have to exchange the books which interest me for textbooks’
 TH: Knihy, ktoré ma zaujímajú, som musela vymeniť za učebnice

Delayed placement occurs in cases when attributive subordinate clause splits the main clause so that the initial noun phrase with subordinate clause behaves as ‘bad clitic host’. Pancheva (2005) explains this pattern as a consequence of some need for clitics to be ‘second’ in their I-phrase.

The term ‘prosodic host’ should be differentiated from the term ‘anchor’, which is also used with respect to clitic components. According to Haspelmath (2022), anchor is the word preceding an enclitic, and the word following a proclitic, whereas host is the element with which a clitic forms prosodic word. Slovak clitics are prosodically neutral which means that they can lean prosodically to the following or preceding word depending on several factors. Prosodically, it is often determined by the length of the anchor. However, although prosodically neutral, clitics cannot follow just any pause but there are also syntactic factors relevant to the position of clitics. In those syntactic structures in which the subordinate clause precedes the main

clause, the clitic is placed immediately after the I-phrase boundary only if the subordinate clause fulfils the role of the relative content clause:

- (14) Čo to bolo, mi nepovedal. (OSIVM)
 what it-NOM be-PST-NEUTR.SG me-DAT NEG-tell-MASC.SG
 ‘What it was he did not tell me’

The possible explanation for clitic ordering in main clause (*mi nepovedal*) is that the object (*čo*) of the complement predicate (*nepovedal*) is extracted from the main sentence and is realized as a complementizer introducing the subordinate clause, thus representing the first syntactic phrase after which the clitic component is placed in the expected 2P position. However, when no such extraction occurs, the subordinate clause is not parsed as an anchor for clitic and the clitic component thus occurs after the initial syntactic phrase of the main clause, as in example (18).

- (15) Keď sa vrátil, podal mi ruku (OSIVM)
 when REFL come_back-PST-MASC.SG give-PST-MASC.SG me-DAT hand-ACC
 ‘When he came back, he gave me his hand’

The identification of subordinate clauses with extraction and without extraction requires a high level of syntactic competence from the learners. The data show that the most common error in texts is the placement of clitic components immediately after the I-phrase boundary even in clauses with no extraction, as in example (19):

- (16) Keď som už chodila do siedmeho
 when be-1SG already attend-PST-FEM.SG in seventh
 ročníka, som dostala svoj prvý
 grade-GEN be-1SG get-PST-FEM.SG my first
 mobilný telefón. (hu, B2)
 mobile phone-ACC
 ‘When I was attending the seventh grade, I got the first mobile phone’
 TH: Keď som už chodila do siedmeho ročníka, dostala som svoj prvý mobilný telefón

5.3 Errors with respect to distance of clitic component from its matrix verb and proficiency levels

The next parameter being analysed is the expected proximity of the clitic component from its matrix verb in terms of units labelled as linear dependency segments (LDS). Mačutek et al. (2021, 331) define the LDS as follows: “[...] the longest possible sequence of words (belonging to the same clause) in which all linear neighbours (i.e. words adjacent in a sentence) are also syntactic neighbours [...]”. It can be illustrated by the sentence from the data:

- (17) Tisícročné tradície dnes sa už
 Millenial traditions-NOM today REFL already
 menej a menej odrážajú v najmladšej generácii. (uk, B2)
 less and less reflect-PRS-3PL in youngest generation-LOC
 ‘Thousand years old traditions are already reflected in the youngest generation a little less’
 TH: Tisícročné tradície sa dnes už menej a menej odrážajú v najmladšej generácii

Four indices are calculated in the data annotation:

- (i) Expected distance between clitic and its matrix verb measured in the number of LDS (ED_{LDS}) and actual distance between clitic and its matrix verb measured in the number of LDS (AD_{LDS});
- (ii) Expected distance between the clitic and its matrix verb measured in the number of syntactic words (ED_{SW}) and the actual distance between the clitic and its matrix verb measured in the number of syntactic words (AD_{SW}).

In example (20), the expected distance of clitic component and its matrix verb with respect to number of LDS is 3 (TH: *Tisícročné tradície sa [dnes] [už] [menej a menej odrážajú] v najmladšej generácii.*), the actual distance is 2 (*Tisícročné tradície [dnes] sa [už] [menej a menej odrážajú] v najmladšej generácii.*), the expected distance from the matrix verb with respect to number of syntactic words is 4 (= *dnes, už, menej, menej*), and the actual distance from the matrix verb with respect to number of syntactic words is 3 (*už, menej, menej*).

Discontinuous placement of the clitic component mostly relates to the ability of the learner to move it to the left sentence periphery independently of the position of the matrix verb. In the following figure, the expected and actual distance ratio is calculated, taking into consideration only the left sentence edge. Five distance values were taken into account: expected distance 0 LDS and actual distance 0 LDS (labelled as 0_0, the error is not caused by wrong distance placement in this case), expected distance 1 LDS and actual distance 0 LDS (labelled as 1_0), expected distance 2 LDS and actual distance less than 2 LDS (1 or 0, labelled as 2_10), expected distance 3 LDS and actual distance less than 3 LDS (2, 1 or 0, labelled as 3_210), and expected distance 4 LDS and actual distance less than 4 LDS (3, 2, 1, or 0, labelled as 4_3210).

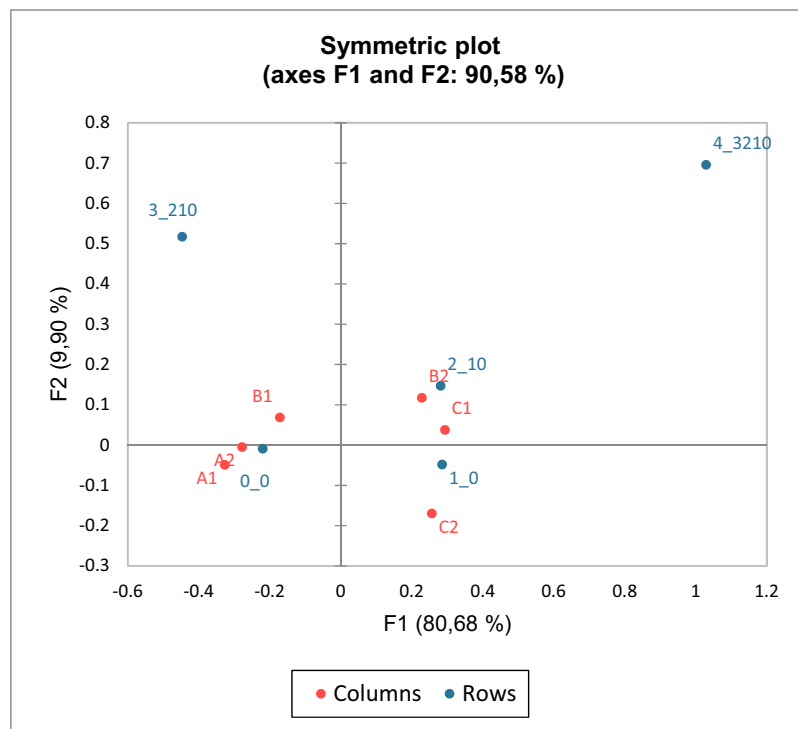


Figure 3: Overall distribution of distance of clitic components from matrix verbs with respect to proficiency levels.

We can summarize the interpretation of the plot as follows:

- A1 and A2 are mostly attracted to 0_0 distance (i.e. the syntactic structure of the sentence does not require non-adjacent placement);
- the distance 3_210 is grouped with B1;
- the distances 2_10 and 4_321 are distinctively associated with B2 and C1.

The data show that the growing syntactic complexity of the sentence structure calls for the distant placement of the clitic component on the left sentence periphery which often results in erroneous placement even at higher proficiency levels (B1, B2, and C1):

- (18) Áno, dieťa má svoje právo vyteru
 yes child-NOM has-PRS-3SG its right-ACC appearance-GEN
 a názoru, len niekedy treba len správne
 and opinion-GEN only sometimes necessary only properly
 nasmerovať ho. (uk, B2)
 direct-INF it-ACC
 ‘Yes, the child has the right to his/her own appearance and opinion, but sometimes it is necessary to direct him/her’
 TH: Áno, dieťa má svoje právo výzoru a názoru, len ho niekedy treba len správne nasmerovať

In example (18), the pronominal clitics *ho* is expected to be realized 4 LDS from its matrix verb (*niekedy + treba + len + správne nasmerovať*); however, its actual distance is 0 LDS.

5.4 Errors with respect to clitic placement on right or left periphery of the matrix verb

Following Zimmerling and Kosta (2013) and Zimmerling (2006), Slavic word-order systems can be classified into four types, tagged as ‘W’ for standard Wackernagel system with consistent 2P-clitics (Serbian, Croatian, Slovenian within Slavic languages in our data); ‘W⁺’ for modified Wackernagel system with verb adjacent clitics (Bulgarian within Slavic languages in our data); ‘W^{*}’ for degraded Wackernagel system with inconsistent 2P-clitics (Polish within Slavic languages in our data); and ‘C’ for communicative system (Russian, Ukrainian within Slavic languages in our data; Zimmerling and Kosta 2013, 201–203).

According to Franks and Holloway King (2000, 61), Bulgarian clitics are verb-adjacent rather than Wackernagel position clitics. However, pronominal clitics in Bulgarian behave comparably to other South Slavic and West Slavic languages, only higher Bulgarian clitics exhibit some idiosyncratic properties, especially clitic component *li*, which does not have an equivalent in Slovak. Therefore, Bulgarian is analysed as an instance of extended Wackernagel clitics system in this article. Similar behaviour is typical of Macedonian clitics; however, Macedonian learners of Slovak as L2 are not included in our data. East Slavic languages lack the kinds of pronominal and auxiliary clitics and the reflexive morpheme *-sja* became frozen as such (Franks and Holloway King 2000, 187).

On the basis of corpus data, the difference between Slavic languages belonging to standard, extended, and degraded W-systems and C-systems with respect to clitic placement on right or left periphery of the matrix verb is calculated. Three position variations are taken into consideration: 1) cases when the expected position of the clitic component is preverbal, yet the realized position is postverbal (PrePost); 2) cases when the expected position of the clitic component is postverbal, yet the realized position is preverbal (PostPre); and 3) cases when both expected and realized positions are preverbal (PrePre). In the latter case, the error arises as a consequence of the non-adequate distance of the clitic component from its matrix verb, not due to position on the left or the right periphery of the matrix work. The case when both expected and realized position is postverbal (PostPost) are not taken into consideration as there are only six instances of this type in the data.

Table 3 brings the percentual distribution of data in the group of L1 Slavic languages with W-systems and Slavic languages with C-systems.

Table 3: Ratio of expected to actual placement of clitic components on the left or right periphery of the matrix verbs in the group of L1 Slavic languages with W-systems and Slavic languages with C-systems

	Slavic languages with W-systems	Slavic languages with C-systems
PrePost	40	45
PostPre	17	10
PrePre	43	45

The differences between frequency distribution within the group of Slavic learners from W-systems and C-systems were statistically tested by CA. The data show that there is no link between the rows and the columns of the table (as the computed p -value 0.341 is greater than the significance level $\alpha = 0.05$, one cannot reject the null hypothesis H_0 : The rows and the columns of the table are independent). The lack of a significant difference between the two groups of Slavic learners is noteworthy. It suggests that Slavic learners, irrespectively of L1 classification of clitics into W-systems or C-systems, place the clitic components on left or right periphery of the matrix verb with a comparable error rate.

Therefore, the next parameter tested is the relation between the positional realization of clitics and the proficiency level. Figure 4 displays the relative positions of both the row vectors – clitic placement on the left or right periphery of the matrix verb – and the column vectors – the proficiency levels.

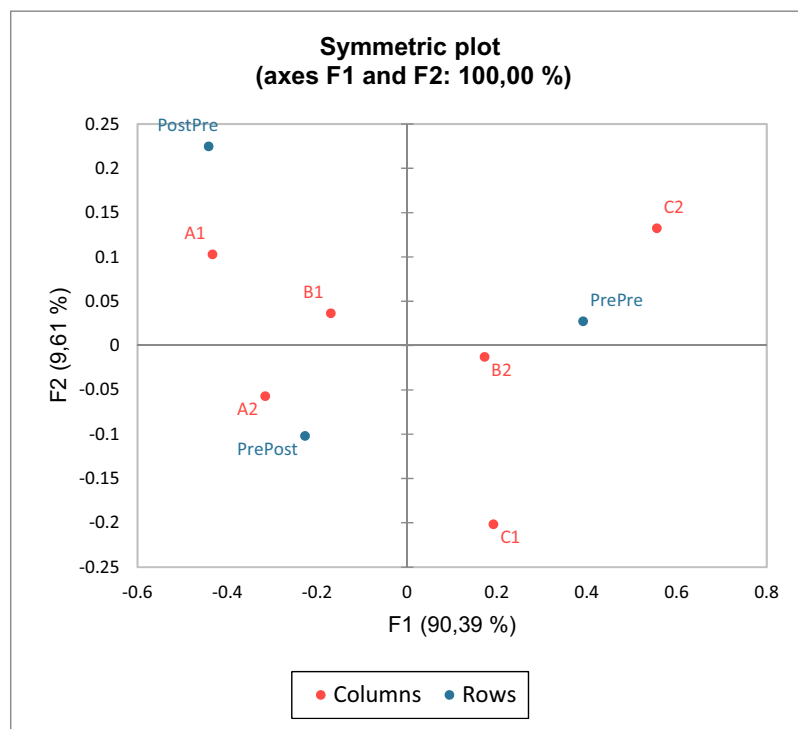


Figure 4: Overall distribution of types of clitic placement with respect to proficiency levels.

In the case of clitic placement with respect to the right or left periphery of the matrix verb, the data show significant differences between the proficiency levels. These findings suggest that clitic placement with respect to matrix verbs undergoes syntactic development.

We can summarize the interpretation of the plot as follows:

- PostPre position is grouped with B1;
- PrePost realization is associated with A2;
- PrePre shows attraction to C2.

The development of word-order competence shows different development trajectories depending on the proficiency levels. Early stages (mostly A2) are typical of erroneous placement of clitic components into postverbal position, i.e. on the right periphery of the matrix verb (PrePost), irrespective of the prosodic pattern of the utterance which calls for placement of clitics on the left periphery of the matrix verb.

- (19) Na ostrov išli sme a videli sme opicy. (hu, A2)
 on island-acc go-PST-PL be-1PL and see-PST-PL be-1PL monkeys-PL-ACC
 ‘We went to the island and we could see the monkeys’
 TH: Na ostrov sme išli a videli sme opice

Another possible explanation of erroneous clitic placement in the first sentence is the role of parallelism: first, grammatical parallelism (i.e. clitic components in a compound sentence are both the auxiliary clitics), and second, positional parallelism, with respect to the position (the position of clitic in the second sentence influences the placement of clitic on the right periphery of the matrix verb in the first sentence).

The following stage (mostly B1) is characteristic of errors concerning placement of clitics into preverbal position instead of postverbal position (PostPre). It can be explained on the basis of development of syntactic complexity in the interlanguage. Numerous theoretical studies point to the fact that syntactic complexity in L2 is thought to expand from coordination to subordination and then to phrasal elaboration, as learners gain proficiency (Wolfe-Quintero et al. 1998). At beginner and low–intermediate proficiency levels, syntactic growth may show an increase in coordination (Bardovi-Harlig 1992, Vyatkina 2012) and upper–intermediate levels are thought to display an increase in subordinate structures. The same results have been proved for Slovak as L2 (Ivanová et al. 2021). Erroneous placement of clitic components into preverbal position occurs mostly in compound sentences whose ratio ascends at A2 and B1 proficiency levels.

- (20) Budúci deň, Soňa kúpila ovocnú
 next day-acc Soňa-NOM buy-PST-FEM.SG fruit
 zmrzlinu a sa prechádzala po pláži
 ice cream-acc and REFL walk-PST-FEM.SG on beach-LOC
 (it, B1)
 ‘Next day Soňa bought herself a fruit ice cream and she took the walk along the beach’
 TH: Nasledujúci deň si Soňa kúpila ovocnú zmrzlinu a prechádzala sa po pláži

In Slovak, elements such as *a* ‘and’ cannot serve as prosodic hosts. The same holds for Bulgarian *i* ‘and’ which nonetheless obviate the reordering phenomenon, e.g.

- (21) Капитанът хвана Паркхил и му
 captain-NOM catch-PST-SG Parkhill-ACC and him-DAT
 изби зъбите. (InterCorp v15 Bulgarian)
 knocked out-3SG teeth-ACC
 ‘The captain caught Parkhill and knocked his teeth out’

In principle, clitics in Bulgarian are preverbal elements. When there is no nonverbal lexical material that could support clitic placement in front of the verb, the verb itself must precede the clitics to provide phonological support for them. This reordering strategy is known as the Tobbler-Mussafia effect. However, even *i* ‘and’ which cannot serve as a prosodic host is sufficient to serve this purpose to keep the clitic component in preverbal position and thus obviate TM to occur (Franks 2008, 99).

In the data, 50% of the errors occurring in this type of syntactic construction was produced by Bulgarian speakers, e.g.

- (22) On kráčal oproti mne a ma
 he walk-PST-MASC.SG towards me-DAT and me-ACC
 pobockal. (bg, A2)
 kiss-PST-MASC.SG
 'He walked towards me and he kissed me'
 TH: On kráčal oproti mne a pobozkal ma

Finally, the last stage of interlanguage development (advanced proficiency levels, mostly C1 and C2) is connected with erroneous placement of clitics concerning their distance from the matrix verb (PrePre). At this stage, learners are able to put the clitic component in the position on the left periphery of the matrix verb, depending on prosodic circumstances in the sentence; however, the errors mostly occur as the result of inappropriate distance of the clitic component from the matrix verb.

The next parameter that was tested was the relation between the positional realization of clitics and the type of clitic component. Figure 5 displays the relative positions of both the row vectors – clitic placement on the left or right periphery of the matrix verb – and the column vectors – the type of clitic component.

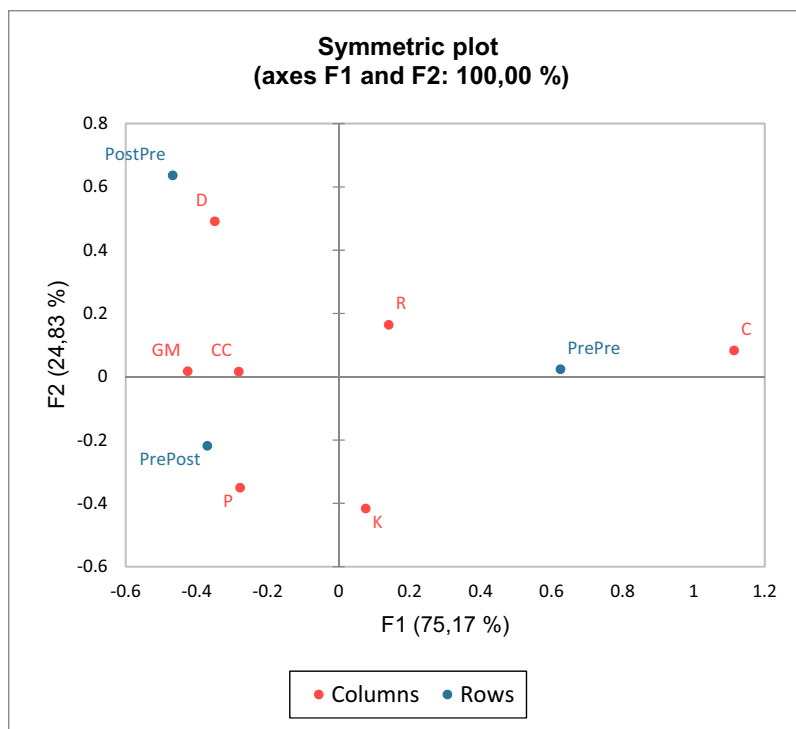


Figure 5: Overall distribution of clitic placement on the left or right periphery with respect to type of clitic component.

In the case of clitic placement with respect to the right or left periphery of the matrix verb, the data show significant differences between the types of clitic components. The CA provides two significant attractions:

- PrePre position is grouped with copular clitic;
- PostPre realization is distinctively associated with demonstrative clitic.

The data show a strong preference for adjacent realization of clitic and its predicative complement in the interlanguage of learners. The expected distance of clitic *be*-form from its predicative complement is 1.06 LDSs and 1.6 syntactic words, yet the actual distance ratio is 0.06 LDSs and 0.07 syntactic words which means that the learners tend to place clitic *be*-form adjacently to its predicative complement irrespective of its clitic nature.

The possible explanation for errors in usage clitic *be*-forms is the nature of the element which occupies the 2P position instead of clitic. According to the data, this element is predominantly subject (49%) or adverbial (27%), i.e. the components which fulfil the role of thematic elements, namely the role of Bearer of Quality (B) as in example (23) or (locational or temporal) Setting (Set) as in example (24):

- (23) Ale väčšinou domáci rozpočet financii
 however mostly domestic budget-NOM finances-GEN
 je postavený na každom zarabajúcom člene
 be-3SG based on every earning member-LOC
 rodiny. (uk, B2)
 family-GEN
 ‘However, in most cases the domestic finance budget is based on every earning member of family’
 TH: Ale väčšinou je domáci rozpočet financií postavený na každom zarabajúcom členovi domácnosti
- (24) Podľa mňa dnes je ľahké sa stretnúť a
 according me-GEN today be-3SG easy REFL meet-INF and
 zoznámiť s novými ľuďmi. (hu, B1)
 get to with new people-INST
 know-INF
 ‘In my opinion, it is easy to meet and get to know new people in these days’
 TH: Podľa mňa je dnes ľahké stretnúť nových ľudí a zoznámiť sa s nimi

In these cases, the clitic *be*-form undergoes delayed placement which is caused by a communicative barrier (forwarding theme element).

Cognitively, the preferential placement of *be*-form in an adjacent position to its predicative complement can be understood as a trade-off between information marking and word order of clitics. It is connected to the idea that communication is a noisy phenomenon and communicators strive to minimize the potential for errors on the side of receivers (noisy channel account; Gibson et al. 2013). In *be*-constructions, both predicative and subject nouns can be used in nominative case in Slovak.

- (25) Ak homosexuál je veriaci, (.) prečo by kostol
 if homosexual-NOM be-3SG believer-NOM why COND church-NOM
 má mu to zakázať. (ru, C1)
 have-3SG him-DAT it-ACC forbid-INF
 ‘If a homosexual is a believer, why the church should forbid it’
 TH: Ak je homosexuál veriaci, (...) prečo by mu to kostol mal zakázať

Within the structure in which subject and predicative noun are located adjacently as in the target hypothesis in example (25) (*ak je homosexual veriaci*), the role difference between subject or predicative noun is lost due to noise which element is the subject and which element is the predicative noun. Interposing clitic components between two nouns is more robust to noise because even when information about one sentence component is cancelled out; the interlocutor can tell solely from the position with respect to the verb whether the information concerns subject or predicate noun: *ak homosexuál je* = subject, *ak je veriaci* = predicate noun. This situation holds for reversible structures (when both subject and predicate noun are semantically homogenous and both components can fulfil the role of subject and predicate noun).

The data also show the preferential placement of demonstratives into preverbal positions. It can be explained by the semi-clitic status of demonstrative components which under certain circumstances may occupy the initial position.

Demonstrative pronoun functions as a clitic component in structures in which it refers to a topicalized antecedent noun phrase. In these constructions, the reference can be viewed as a process of elaboration (the noun phrase is being elaborated on or made more accurate) with the classifying or identificational adjectives modifying predicative complement. The canonical position of demonstrative elements behaving as volatile

clitics is 2P in these types of structures. On the other hand, non-clitic realization is preferred for cases when the clitics resume a whole sentence or text passage and the reference can be seen not as elaborating but as evolving (it signals a new denomination, often expressing also evaluative attitude).

The differentiation of reference as elaborating or evolving requires a high level of pragmatic and syntactic competence from the learners. The data show that the most common error in texts is the placement of demonstrative clitic components into initial position in clauses with elaborating references as in example (26) or cliticization of demonstrative elements in clauses with evolving references as in example (27):

- (26) V Číne najdeme najvyššiu a najmladšiu
 in China-LOC find-1PL highest and youngest
 plošinu. To je Tibetská plošina. (zh, A1)
 plateau-ACC It-NOM be-3SG Tibet's plateau-NOM.
 'In China one can find the highest and youngest plateau. It is Tibet's plateau'
 TH: V Číne nájdeme najvyššiu a najmladšiu plošinu. Je to Tibetská plošina

- (27) Ako peniaze vplyvajú na správanie človeka
 how money-NOM impact-PRS-3PL on behaviour-ACCman-GEN
 a či kazia jeho charakter? Je
 and whether spoil-PRS-3PL his character-ACCbe-3SG
 to celkom zaujímavá otázka. (pl, B1)
 it-NOM quite interesting question-NOM
 'How does money influence the human behaviour and does it spoil his character?'
 TH: Ako peniaze vplyvajú na správanie človeka a kazdia jeho charakter? To je celkom zaujímavá otázka

Presented word-order patterns differ with respect to index referred to as 'wdist' (distance between anaphor and antecedent in words; Strube et al. 2002), which is usually described as one of the coreference-level features contributing to successful anaphor resolution (linking the referring expression to its antecedent). Within Accessibility Theory (Ariel 2001) anaphor resolution is linked to the appropriate degree of cognitive accessibility of the antecedent referent. Accessibility depends on several factors: the antecedent's salience, its recency of mention, and cohesion between clauses containing antecedent and anaphor, etc. Clitic usage of demonstrative pronoun posits wdist 1 in clauses with elaborating reference, whereas its non-clitic usage leads to wdist 0 in clauses with evolving reference.

In example (26), the antecedent noun phrase *najvyššiu a najmladšiu plošinu* has a high level of salience – it has been shown that noun phrases containing attributes as well as referents being the goal of action (*nájdeme plošinu*) are perceived as more prominent than noun phrases without attributes or with other semantic roles (Klin et al. 2004, Arnold 2001). High-salience antecedents can have their referring expressions (i.e. demonstrative clitic) more distantly than low-salience referring expressions.

On the other hand, the demonstrative element in example (27) resumes the whole previous sentence passage which as an anaphoric non-nominal antecedent has a lower degree of salience. This fact is thus balanced by zero distance of anaphoric demonstrative from its referring expression, resulting into its non-clitic status. Thus, each position realization of demonstrative can be interpreted as referring to a referent of a particular salience level. Non-clitic use of a demonstrative signals a less salient antecedent referent whereas a more salient antecedent referent admits clitic usage of demonstrative.

6 Discussion and conclusions

The study of word order of clitics in the interlanguage of students acquiring Slovak as L2 contributes to investigations into second-language acquisition, cognitive linguistics, and Slovak linguistics. It lends empirical evidence to the important question raised in the theory of interlanguage, namely if interlanguage and its

development are governed by some general cognitive principles which could explain the preferential uses of certain word-order patterns in the interlanguage of learners, irrespective of their L1.

The word order of clitics is a complex phenomenon that has received considerable attention in Slavic linguistics, but no empirical investigation of contemporary corpus data has been carried out for Slovak as L2. The present study fills this gap by offering a statistical analysis of a database of more than 1,000 examples. At the beginning of the study, three research questions were formulated. Our contribution to the study of Slovak clitics can be summarized as follows. Learners of Slovak as L2 prefer those constituent orders:

- (i) which maintain adjacency of clitic components and their matrix verbs (in accordance with the contiguity principle),
- (ii) in which clitic elements are placed on the right periphery of their matrix verbs (in accordance with the principle of relevance),
- (iii) in which argument clitics follow the matrix verb, they overrely on SVO orders (in accordance with the principle of verb–object bonding).

These data are sufficient to stipulate the basic properties of clitics in the interlanguage of Slovak: they are preferentially realized as verb-adjacent and postposed elements. At the same time, the data proved that the tendency for adjacent placement of clitic component on the right periphery of the matrix verb (Matrix Verb = C) is not skewed by the L1 of the students (the data were tested for Slavic languages with identical or similar and disparate clitic systems with no statistically relevant deviation). It can be assumed that where there is a common systematic preference for some word order over another across language groups with different characteristics, then the underlying cause must primarily be cognitive-functional, not interferential.

What has been proved as a statistically relevant parameter is the relation between clitic placement and proficiency levels (Figures 3 and 4), which points to the existence of certain developmental patterns in word order competence concerning clitics placement in the interlanguage of the learners. These results prove that the word order of clitics seems to be immune to direct influence from L1 in most cases (the only exception is the frequent placement of clitics after certain connectives which cannot serve as prosodic hosts in texts produced by Bulgarian learners), and a more important role is played by general cognitively anchored principles which are at stake in different stages of interlanguage development.

At the same time, the research on clitic ordering makes an important contribution to the question of how well L2 learners are able to integrate linguistic phenomena pertaining to interfaces. Word order of clitics is partly located at external interfaces (prosody–syntax–information structure), which proves to be a locus of instability even at higher proficiency levels – the data show that there is a clear tendency for speakers to produce orders with delayed placement of clitic components due to syntactic barriers (when 2P position is occupied by subject argument, subject-first preference) or communicative barriers (when 2P position is occupied by the topic element, topic-first preference). The findings confirm the idea of the emergentist usage-based view of interlanguage grammar which builds in developmental stages in the form of sequential associations of different strengths being weighted and implemented in language production, depending on the degree of entrenchment of a sequence in an individual mind.

Funding information: This work was supported by the Slovak Research and Development Agency within contract No. APVV-19-0155 Language Errors in Slovak as a Foreign Language Based on Learner Corpus (80%) and No. KEGA 029PU-4/2022 Slovak as a Second Language Online (with a Primary Focus on Ukrainian Students) (20%).

Conflict of interest: The author states no conflict of interest.

Data availability statement: All data (and related metadata) are publicly available via the Zenodo platform at <https://doi.org/10.5281/zenodo.8332408> (Ivanová 2023).

References

- Ariel, Mira. 2001. "Accessibility theory: An overview." In *Text representation: Linguistic and psycholinguistic aspects*, edited by Ted J. M. Sanders, Joost Schliperoord, and Wilbert Spooren, p. 29–87. Amsterdam, Philadelphia: John Benjamins.
- Arnold, Jennifer E. 2001. "The effects of thematic roles on pronoun use and frequency of reference." *Discourse Processes* 31(2), 137–62.
- Avgustinova, Tania and Karel Oliva. 1995. "On the nature of the Wackernagel position in Czech." In *Formale Slavistik*, edited by Uwe Junghanns and Gerhild Zybatow, p. 25–47. Frankfurt am Main: Vervuert.
- Bardovi-Harlig, Kathleen. 1992. "A second look at T-unit analysis: Reconsidering the sentence." *TESOL Quarterly* 26, 390–5.
- Bates, Elizabeth and Brian MacWhinney. 1982. "Functionalism approaches to grammar." In *Language acquisition: The state of the art*, edited by Eric Wanner and Lila R. Gleitman, p. 173–218. Cambridge: Cambridge University Press.
- Bates, Elizabeth and Brian MacWhinney. 1989. "Functionalism and the competition model." In *The crosslinguistic study of sentence processing*, edited by Elizabeth Bates and Brian MacWhinney, p. 3–76. New York: Cambridge University Press.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan L., William Pagliuca, and Revere D. Perkins. 1990. "On the asymmetries in the affixation of grammatical material." In *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*, edited by William Croft, Keith Denning, and Suzanne Kemmer, p. 1–42. Amsterdam: John Benjamins.
- Claesen, Harald and Pieter Muysken. 1989. "The UG paradox in L2 acquisition." *Interlanguage Studies Bulletin (Utrecht)* 5(1), 1–29.
- Clancy, Steven. 2010. *The Chain of Being and Having in Slavic (= Studies in Language Companion Series 122)*. Amsterdam, Philadelphia: John Benjamins.
- Cho, Jacee and Roumyana Slabakova. 2014. "Interpreting definiteness in a second language without articles: The case of L2 Russian." *Second Language Research* 30(2), 159–90.
- Dressler, Wolfgang U., Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel. 1987. *Leitmotive in natural morphology*. Amsterdam, Philadelphia: John Benjamins.
- Dryer, Matthew S. 2013. "Order of subject, object and verb." In *WALS Online. (v2020.3) [Data set]*. Zenodo, edited by Matthew S. Dryer and Martin Haspelmath. doi: 10.5281/zenodo.7385533.
- Ellis, Rod. 2008. *The study of second language acquisition*. Oxford: Oxford University Press.
- Franks, Steven. 2008. "Clitic placement, prosody, and the bulgarian verbal complex." *Journal of Slavic Linguistics* 16(1), 91–7.
- Franks, Steven and Tracey Holloway King. 2000. *A handbook of slavic clitics*. Oxford: Oxford University Press.
- Gehrke, Berit. 2018. "Cross-Slavic variation in passives." *Workshop on Aspect and Argument Structure of Adverbs/Adjectives and Prepositions/Participles (WAASAP 4)*. Universitat Pompeu Fabra, Barcelona, June 7–8, 2018.
- Gell-Mann, Murray and Merritt Ruhlen. 2011. "The origin and evolution of word order." *Proceedings of the National Academy of Science of the United States of America* 108(42), 17290–5.
- Gibson, Edward. 1998. "Linguistic complexity: Locality of syntactic dependencies." *Cognition* 68(1), 1–76.
- Gibson, Edward, Leon Bergen, and Steven T. Piantadosi. 2013. "Rational integration of noisy evidence and prior semantic expectations in sentence interpretation." *Proceedings of the National Academy of Sciences* 110(20), 8051–6.
- Glynn, Dylan. 2014. "Correspondence analysis." *Methods for Semantics: Quantitative Studies in Polysemy and Synonymy* 43, 443–85.
- Hahn, Michael, Dand Jurafsky, and Richard Futrell. 2020. "Universals of word order reflect optimization of grammars for efficient communication." *Proceedings of the National Academy of Sciences of the United States of America* 117(5), 2347–53.
- Hall, Matthew L., Rachel I. Mayberry, and Victor S. Ferreira. 2013. "Cognitive constraints on constituent order: Evidence from elicited pantomime." *Cognition* 129(1), 1–17.
- Hana, Jirka. 2007. *Czech clitics in higher order grammar*. Columbus: The Ohio State University.
- Haspelmath, Martin. 1993. "The diachronic externalization of inflection." *Linguistics* 31(2), 279–9.
- Haspelmath, Martin. 2022. "Types of clitics in the world's languages." *Linguistic Typology at the Crossroads*, 3(2), 1–59.
- Haspelmath, Martin. 2023. "Is cliticization an intermediate stage between free lexeme and affix status?" *25th International Conference on Historical Linguistics (ICHL 25)*, Oxford, Zenodo. doi: 10.5281/zenodo.7054292.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hulk, Aafke. 1991. "Parameter setting and the acquisition of word order in L2 French." *Second Language Research* 7(1), 1–34.
- Ivanová, Martina. 2023. Dataset for the paper "Under pressure. Exploring the impact of cognitive factors on clitics placement in L2 Slovak" [Data set]. Zenodo. doi: 10.5281/zenodo.8332408.
- Ivanová, Martina, Miroslava Kyseľová, and Anna Gálisová. 2021. "Acquiring word order in Slovak as a foreign language: Comparison of Slavic and Non-Slavic learners utilizing corpus data." *Jazykovedný Časopis* 72(2), 353–70.
- Jiang, Wenying. 2009. *Acquisition of word order in chinese as a foreign language*. Berlin: Mouton de Gruyter.
- Junghanns, Uwe. 2002. "Clitic climbing im Tschechischen." In *Untersuchungen zur Syntax und Informationsstruktur slavischer Deklarativsätze (=Linguistische Arbeitsberichte 80)*, edited by Uwe Junghanns, p. 57–90. Leipzig: Institut für Linguistik der Universität Leipzig.
- Klin, Celia M., Kristin M. Weingartner, Alexandria E. Guzman, and William H. Levine. 2004. "Readers' sensitivity to linguistic cues in narratives: How salience influences anaphor resolution." *Memory and Cognition* 32(3), 511–22.
- Koch, Peter. 1999. "Frame and contiguity. On the cognitive basis of metonymy and certain types of word formation." In *Metonymy in language and thought*, edited by Klaus-Uwe Panther and Günter Radden, p. 139–67. Amsterdam, Philadelphia: John Benjamins.

- Kosek, Pavel. 2011. *Enklitika v češtině barokní doby*. Brno: Host.
- Kisselev, Olesya. 2019. "Word order patterns in the writing of heritage and second language learners of Russian." *Russian Language Journal* 69(1), 149–74.
- Lehmann, Christian. 2002. *Thoughts on grammaticalization – second, revised edition*. Erfurt: Universität Erfurt, Philosophische Fakultät.
- Mačutek, Ján, Radek Čech, and Marine Courtin. 2021. "The Menzerath-Altmann law in syntactic structure revisited." In *Quasy, SyntaxFest 2021: Proceedings of the Second Workshop on Quantitative Syntax (March 21–25, 2022)*, p. 65–73. Sofia: Association for Computational Linguistics.
- Maurits, Luke, Andrew Perfors, and Danielle Navarro. 2010. "Why are some word orders more common than others? A uniform information density account." In *Advances in Neural Information Processing Systems 23 (NIPS 2010)*, p. 1585–93. Cambridge, MA: MIT Press.
- Mistrík, Jozef. 1966. *Slovosled a vetosled v slovenčine*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.
- Ortega, Lourdes. 2009. *Understanding second language acquisition*. London: Hodder.
- Pancheva, Roumyana. 2005. "The rise and fall of second position clitics." *Natural Language and Linguistic Theory* 23(1), 103–67.
- Pienemann, Manfred. 2005. *Cross-linguistic aspects of processability theory*. Amsterdam: John Benjamins.
- Qian, Ting and Florian Jaeger. 2012. "Cue effectiveness in communicatively efficient discourse production." *Cognitive Science* 36(7), 1312–36.
- Rosen, Alexandr. 2001. *A constraint-based approach to dependency syntax applied to some issues of Czech word order*. Ph.D. thesis. Prague: Charles University.
- Sorace, Antonella and Francesca Filiaci. 2006. "Anaphora resolution in near-native speakers of Italian." *Second Language Research* 22(3), 339–68.
- Sorace, Antonella and Ludovica Serratrice. 2009. "Internal and external interfaces in bilingual language development: Beyond structural overlap." *International Journal of Bilingualism* 13(2), 195–10.
- Spencer, Andrew and Ana R. Luis. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press.
- Strube, Michael, Stefan Rapp, and Christoph Müller. 2002. "The influence of minimum edit distance on reference resolution." In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, p. 312–9. Philadelphia: Association for Computational Linguistics.
- Tomaselli, Alessandra and Bonnie D. Schwartz. 1990. "Analysing the acquisition stages of negation in L2 German: support for UG in adult SLA." *Second Language Research* 6(1), 1–38.
- Tomlin, Russel S. 1986. *Basic word order: Functional principles*. London: Routledge.
- Uhlířová, Ludmila, Petr Kosta, and Ludmila Veselovská. 2017. "Klitikon." In *CzechEncy – Nový encyklopedický slovník češtiny*, edited by Petr Karlík, Marek Nekula, and Jana Pleskalová. Brno: Masarykova univerzita. <https://www.czechency.org/slovník/KLITIKON>.
- Vyatkina, Nina. 2012. "The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study." *Modern Language Journal* 96(4), 576–98.
- Wolfe-Quintero, Kate, Shinji Inagaki, and Hae-Young Kim. 1998. *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Centre.
- Zimmerling, Anton W. 2006. "Encoding strategies in word order: the evidence of Slavic languages." In *The 1st meeting of the Slavic Linguistic Society*. Bloomington, 8–10 September, 2006.
- Zimmerling, Anton W. 2008. "Clitic particles and the typology of 2P languages." *Paper presented at the Particle Workshop*, Downing College, Cambridge. <https://antonzimmerling.files.wordpress.com/2009/04/cliticpartzimmerling091.pdf>.
- Zimmerling, Anton W. 2015. "Parametrizing verb second languages and clitic second languages." In *WORLDCOMP' 15. Proceedings of the 2015th International Conference on Artificial Intelligence*, Vol. 1, p. 281–7. Las Vegas: CSREA Press.
- Zimmerling, Anton W. and Petr Kosta. 2013. "Slavic clitics: A typology." *Language Typology and Universals. Sprachtypologie und Universalienforschung* 66(2), 178–14.

Sources

- Czech National Corpus – InterCorp (version 15)*. Institute of the Czech National Corpus. Accessed September 6, 2023. <https://www.korpus.cz/kontext>.
- Errkorp-1.0*. Jazykovedný ústav Ľ. Štúra SAV. Accessed August 16, 2023. <https://bonito.korpus.sk/>.
- Omnia Slovaca IV Maior*. Jazykovedný ústav Ľ. Štúra SAV. Accessed August 4, 2023. <http://ske.juls.savba.sk/bonito/index.html>.