

John Benjamins Publishing Company



This is a contribution from *Language and Text. Data, models, information and applications*.

Edited by Adam Pawłowski, Jan Mačutek, Sheila Embleton and George Mikros.

© 2021. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

On the impact of the initial phrase length on the position of enclitics in Old Czech

Radek Čech¹, Pavel Kosek², Olga Navrátilová²
and Ján Mačutek^{3,4}

¹University of Ostrava / ²Masaryk University / ³Mathematical Institute of Slovak Academy of Sciences / ⁴Constantine the Philosopher University in Nitra

This paper presents an analysis of the relationship between the length of the initial phrase and the positions of pronominal enclitics in a clause. The hypothesis predicting the negative correlation between the length of the phrase and the proportion of enclitics in the post-initial position was set up and tested. For testing the hypothesis, selected books – Genesis (Gn), Isaiah (Is), Job (Jb), Sirach (Sir), Gospel of St. Matthew (Mt), Gospel of St. Luke (Lk), Acts (Act), and Revelation (Rev) – from the first edition of the Old Czech Bible translation were used. The hypothesis was not rejected; however, some differences among particular pronouns were revealed.

Keywords: enclitics, word order, initial phrase, length, Old Czech

1. Introduction

Enclitics are language units with specific phonological as well as syntactic behaviour. They are defined as unstressed words that are joined prosodically with the preceding word. However, a syntactic relationship between the enclitic and that preceding word is not necessary. These facts have a crucial impact on their word order position. Further, several classes of enclitics are determined in linguistics (e.g., auxiliary, pronominal, clausal) and each type displays some specificity (cf. Uhlířová et al. 2017). Moreover, the behaviour of an enclitic depends also on a character of its host (the word to which it is attached). For instance, one can find different word order characteristics of Czech enclitics joined to infinitive verbs, on the one hand, and to finite verbs, on the other (Toman 2000). Consequently, enclitics are considered a heterogeneous set of units which, however, share some common qualities

(Zwicky 1994). One of the possibilities for a better understanding of the properties of enclitics is an analysis of their diachronic development. It allows us to observe forces and mechanisms which have an impact on them and, subsequently, it could bring new ways of how to explain their status in a synchronic language system.

In this paper, we introduce the results of an analysis of the most frequent pronominal enclitics in Old Czech, namely, the enclitics *sě* (accusative reflexive), *mi* (“to me”), and *tě* (“you”). In the investigated texts, the reflexive forms *sě* and *tě* preserve some remnants of stressed (orthotonic) forms. This situation can be viewed as a symptom of the historical change from an unstable enclitic into a stable enclitic / enclitic tantum (Uhlířová et al. 2017). This process probably accelerated during the Old Czech period (Trávníček 1956: 147; Šlosar 1967: 252). As we have shown earlier, in the overwhelming majority of examples the investigated forms are characterized by the properties of an enclitic word (Kosek et al. 2018a). For illustration, we give examples of typical usage of *sě* (accusative reflexive), *mi* (“to me”), and *tě* (“you”) in Old Czech in (1), (2), and (3), respectively (the enclitic is underlined).

- (1) *Kde jest ten, [jenž sě jest narodil,]*
 Who.LOC.F.SG REFL.ACC be.AUX.PRS.3SG born.PTCP.PST.M.SG
král židovský?
 king.NOM.M.SG Jewish.NOM.M.SG
 “Where is the one who has been born king of the Jews?”
 The Olomouc Bible, Mt 2,2
- (2) *nepodal-s mi vody*
 Neg-give.PTCP.PRET.M.SG-be.AUX.PRS.2.SG me.DAT.SG water.GEN.F.SG
mým nohám
 my.DAT.F.PL foot.DAT.F.PL
 “You gave me no water for my feet” The Olomouc Bible, Lk 7,44
- (3) *A v prosbě za Izmahele sem tě*
 and in prayer.LOC.F.SG for Ishmael.ACC.M.SG be.AUX.PRS.1.SG you.ACC.SG
uslyšal
 hear.PTCP.PRET.M.SG
 “As for Ishmael, I have heard you” The Olomouc Bible, Gen 17,20

In this study, we set up a hypothesis considering a relationship between the length of the initial syntactic phrase in a clause and the occurrence (or its absence) of the enclitic after this phrase. The theoretical background of the hypothesis – namely, the longer the first phrase of the clause, the lower the probability of the occurrence of the enclitic after this phrase – puts together both prosodic and syntactic properties of observed phenomena. Specifically, the enclitic cannot appear after a pause. The longer the initial phrase, the higher the probability that a pause will occur after it, and consequently, the probability that the enclitic occurs after the initial phrase

decreases with increasing phrase length. This mechanism is valid only for the initial phrase which does not contain the governor of the enclitic. This phenomenon was described, among others, by Ertl (1924), Radanović-Kocić (1996: 435) who called it a “heavy constituent constraint”, or by Čavar and Wilder 1999: 443–444), who considered it a result of the “clitic third” rule.

For the analysis, data from the oldest Czech Bible translation are used. These data were chosen because this study is one of the first results of a larger project (Kosek et al. 2019) which aims at both describing and explaining word order characteristics of pronominal enclitics and their development from Old Czech to the present. In other words, we start with the oldest language material available in Czech because it represents a ‘starting point’ for our long-term research interest.

2. Classification of the word ordering of enclitics in Old Czech

A detailed descriptive analysis of pronominal enclitics in finite verb phrases reveals several tendencies of its positions in the Old Czech word order (Kosek et al. 2018a). Kosek et al. determine several word order positions of pronominal enclitics and, further, they found significant differences among frequencies of these forms in particular positions. For instance, a pronominal enclitic occurs most frequently after an initial syntactic phrase / the first word of the clause and, on the contrary, it never occurs as the first word of the clause. Two positions are dominant: (1) the post-initial position and (2) the contact position in the middle or at the end of a clause (which we call ‘non-post-initial positions’ for the sake of simplicity). These positions cover about 95% of all occurrences. As for the former, it is defined as the position after the initial phrase, which can be comprised of one or more words, as is illustrated in examples (4) and (5) (square brackets mark the initial phrase, the enclitic is underlined).

- (4) [*Tehda*] sě *otevře* *chrám* *boží*
 then REFL.ACC open.FUT.3SG temple.NOM.M.SG god.ADJ.POS.NOM.M.SG
v nebi
 in heaven.LOC.N.SG
 “Then the temple of God in heaven was opened”
The Olomouc Bible, Rev 11,19
- (5) [*Co*] sě *tobě* *vidí,* *Šimone?*
 what.NOM REFL.ACC you.DAT see.PRAES.3.SG Simon.VOC.SG.M
 “What do you think, Simon?”
The Olomouc Bible, Mt 17,24

As for the non-post-initial position, it is defined as the position in which the enclitic does not occur in the second position and is immediately adjacent to its syntactically/morphologically superordinate item, see example (6), (7), and (8) (squared brackets mark the initial phrase, the superordinate item of the enclitic is bolded, and the enclitic is underlined).

- (6) [*Ale mládenečky*] ***hnětla*** *sě* *v*
 but children.INSTR.M.PL struggle.PART.PRET.ACT.F.SG REFL.ACC in
životě jsúce
 womb.LOC.M.SG being.PART.PRAES.ACT.NOM.PL
 “But the children struggled in her womb” The Olomouc Bible, Gen 25,22
- (7) [*Volánie Sodomských a Gomorrejských*]
 outcry.NOM.N.SG sodom.ADJ.GEN.M.PL and gomorrha.ADJ.GEN.M.PL
rozmnožilo *sě* ***jest***
 multiply.PTCP.PRET.ACT.N.SG REFL.ACC be.AUX.PRET.3SG
 “The cry of Sodom and Gomorrha is multiplied”
 The Olomouc Bible, Gen 18,20
- (8) [*Narozenie*] *pak našeho Jezukrista takto*
 birth.NOM.N.SG then our.GEN.M.SG Jesus.GEN.M.SG Christ.GEN.M.SG so
sě ***jest*** ***stalo***
 REFL.ACC be.AUX.PRET.3SG happen.PTCP.PRET.ACT.N.SG
 “This is how the birth of Jesus the Messiah came about”
 The Olomouc Bible, Mt 1,18

Because the purpose of the study is to analyze the strongest tendencies of word ordering of the enclitics as well as to try to find mechanisms that can explain them, we do not follow the rather fine-grained classification of Kosek et al. (2018a). For the testing of the hypothesis, we determine only two positions: (1) post-initial and (2) non-post-initial (for more details, see § 4).

3. Language material

We used the first edition of the Old Czech Bible translations. We chose the Bible text for the following reasons: (a) it is one of the oldest Old Czech prose texts (older texts from the first half of the 14th century are poetic and they cannot be used to observe word order characteristics typical for the Czech language of that time), (b) the results can be compared with the ones based on later Czech Bible translations which enables us to observe the historical development of the phenomena under study. According to Kvas (1997: 43) and Vintř (2008: 1883a), the complete Old Czech translation of the Bible probably dates from the 1350s and it is considered

to be a work of around ten anonymous translators. However, no autograph of the translation is known. The oldest version of the Old Czech Bible has survived in later copies: the Dresden Bible (*Bible drážďanská*, 1360s), the Litoměřice-Třeboň Bible (*Bible litoměřicko-třeboňská*, 1411–1414), and the Olomouc Bible (*Bible olomoucká*, 1417 – Kyas 1997: 57; Vintr 2008: 1883b). Unfortunately, none of these copies are entirely identical to the original version of the text. Specifically, the original text was slightly revised in later copies and, moreover, some parts of the original have been replaced by more recent translations – e.g., the Litoměřice-Třeboň and Olomouc Bibles include a different translation of the Gospel of Matthew (known as the Gospel of Matthew with homilies), and the Olomouc Bible incorporates some epistles from the Acts of the Apostles, which were taken from the second edition of the Old Czech Bible translation (Kyas 1997: 42, 61–62; Vintr 2008: 1883b). Further, the copies have not survived in their entirety. For instance, the Dresden Bible, which represents the oldest version, was unfortunately entirely destroyed during the First World War, and only part of the original text has survived in the form of photocopies and copies.

Faced with this state of affairs, we chose the Olomouc Bible as the main text for the investigation for the following reasons. It is the oldest known complete text of the Old Czech Bible and it is the text which forms the basis of a critical edition of the Old Czech Bible as conceived by Kyas (Kyas ed. 1981, 1985, 1988; Kyas et al. eds. 1996; Pečírková et al. eds. 2009). Because all observed phenomena had to be annotated manually, we used only a sample of the Bible. We attempted to select books which (1) as far as possible, differ in their text structure and style, and (2) were the work of different translators, in the view of Kyas (1997: 43) who identifies two distinct groups of translators on the basis of the Czech equivalents they used for specific Latin words such as *adorare*, *benedicere*, *benedictus*, etc. In this way, we attempted to at least partially compensate for the limitations inherent in analyzing the specific language of Bible texts. As a result, four books from the Old Testament and four books from the New Testament were chosen: Genesis (Gn), Isaiah (Is), Job (Jb), Sirach (Sir), Gospel of St. Matthew (Mt), Gospel of St. Luke (Lk), Acts (Act), and Revelation (Rev). Since the Acts of the Apostles were taken from the second edition of the Old Czech Bible translation in the Olomouc Bible, this text is taken from the Litoměřice-Třeboň Bible (meaning that the translation is from the first edition of the Old Czech Bible) into our sample.

4. Methodology

To test the hypothesis predicting a relationship between the length of the initial phrase and word ordering of the enclitic, first, we manually determined both the initial phrase and the position of the enclitic in each clause. Next, the length of the initial phrase was measured by the number of (a) letters and (b) words. Finally, for each value of length of the initial phrase, a proportion of the post-initial positions of the enclitic was calculated. However, for some values of length (especially for longer ones) of the initial phrase, we had too few occurrences of enclitics. Therefore, we pooled the data to get at least 5 occurrences of the enclitic in each group and, consequently, we computed the average length of the phrase in such a group. To take the frequency of the occurrence of the enclitic in a particular position into account, we computed average weighted lengths (ALi), where weights are represented by particular frequencies.¹

5. Results

According to the hypothesis, the longer the initial phrase, the smaller the proportion of enclitics in the post-initial position. The length of the initial phrase was measured by the number of letters, see Table 1 and Figure 1–3. For all these enclitics, we can see a general tendency that is in accordance with the hypothesis. However, closer observation also reveals other important findings. First, there are obvious differences between the characteristics of *sě*, on the one hand, and *mi* and *tě*, on the other. Specifically, in the case of the shortest initial phrases, the proportion of the post-initial position of the enclitic *sě* is strikingly smaller than the proportions of *mi* and *tě*. Further, *sě* displays a different shape of distributions of post-initial positions (hereinafter pp) with regard to the ALi in comparison to *mi* and *tě*. In the case of *sě*, a more or less gradual decreasing tendency appears (except for ALi = 1); however, in the case of *mi* and *tě*, one can see no clear tendency up to ALi ≈ 10, then the proportions of post-initial enclitics decrease. To compute the relationship between ALi and pp, Kendall's rank correlation coefficient was used with results as follows. For *sě*, we get $\tau = -0.75$ (p -value < 0.001), for *mi* and *sě*, $\tau = -0.37$ (p -value < 0.115) and $\tau = -0.45$ (p -value < 0.032).

1. The original data are available here: http://cechradek.cz/data/Cech_et_al_On_the_Impact_of_Initial_Phrase_Length_Data.pdf

Table 1. Proportions of post-initial positions (pp) of enclitics *sě*, *mi*, and *tě* in the analyzed books in the Olomouc Bible. The length of the initial phrase is measured by the number of letters. ALi means the average weighted length of the post-initial phrase, fp the frequency of the particular enclitic in post-initial position, fn the frequency of the particular enclitic in non-post-initial position

<i>sě</i>				<i>mi</i>				<i>tě</i>			
ALi	fp	fn	pp	ALi	fp	fn	pp	ALi	fp	fn	pp
2	16	13	0.55	2	25	1	0.96	1.88	14	2	0.88
3	53	27	0.66	3	62	2	0.97	3	25	5	0.83
4	61	32	0.66	4	33	4	0.89	4	24	1	0.96
5	74	38	0.66	5	48	0	1	5	16	4	0.8
6	56	33	0.63	6	26	5	0.84	6	22	4	0.85
7	44	31	0.59	7	21	0	1	7	18	1	0.95
8	27	25	0.52	8	9	1	0.9	8	17	1	0.94
9	27	24	0.53	9	6	2	0.8	9	6	3	0.67
10	17	18	0.49	10	5	0	1	10	6	0	1
11	5	20	0.2	12.17	3	3	0.5	11.83	3	3	0.5
12	3	12	0.2	19.29	1	6	0.14	13.88	2	6	0.25
13	3	18	0.14					16	2	4	0.33
14	0	18	0					20.4	1	3	0.25
15	0	20	0								
16	0	11	0								
17	0	16	0								
18	1	9	0.1								
19	0	5	0								
21.57	0	7	0								
24.4	0	5	0								
39.14	1	6	0.14								

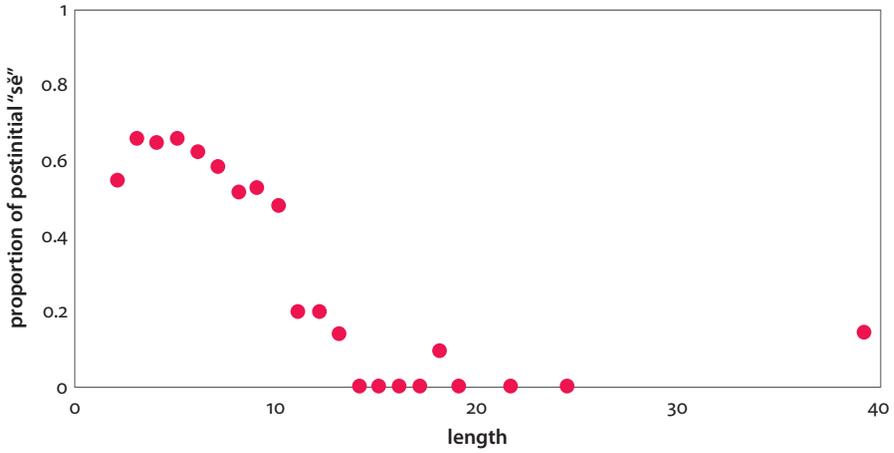


Figure 1. Proportions of post-initial positions of enclitic *sě*. The length of the initial phrase is measured by the number of letters

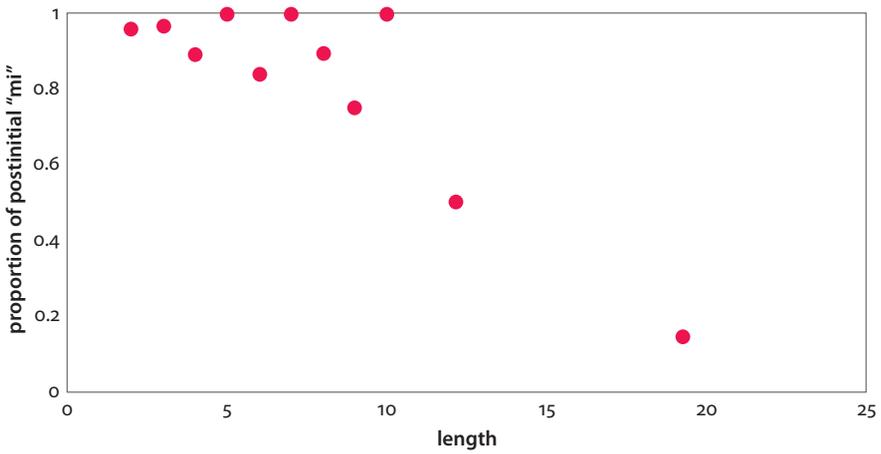


Figure 2. Proportions of post-initial positions of enclitic *mi*. The length of the initial phrase is measured by the number of letters

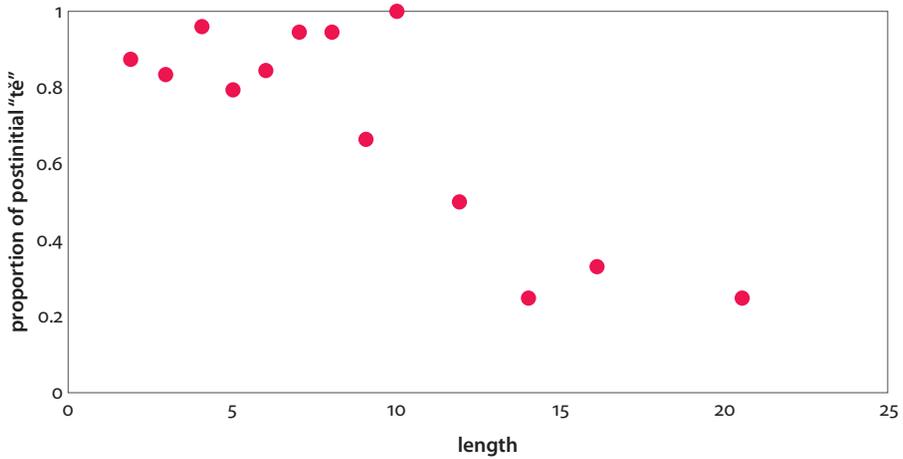


Figure 3. Proportions of post-initial positions of enclitic *tě*. The length of the initial phrase is measured by the number of letters

Alternatively, the length of the initial phrase was measured by the number of words the phrase consists of (see Table 2 and Figure 4–6). For all enclitics, we can see that the general tendency is in accordance with the hypothesis. Further, there is again an obvious difference between the behaviour of *sě*, on the one hand, and *mi* and *tě*, on the other. Namely, the proportions of the post-initial positions of *sě* in the two shortest phrases are smaller in comparison to the respective proportions of *mi* and *tě*. Optically, we can see that the shapes of the distributions differ: *sě* displays a more gradual tendency than *mi* and *tě* which have a minimal difference between ALi = 1 and ALi = 2. However, this optical observation must be taken as very preliminary and only a proper mathematical modelling can reveal if it reflects some tendency or not. In this study, the sample size, unfortunately, does not allow applying either the measurement of a correlation coefficient or of mathematical distributional models.

Table 2. Proportions of post-initial positions (pp) of enclitics *sě*, *mi*, and *tě* in the analyzed books in the Olomouc Bible. The length of the initial phrase is measured by the number of words. ALi means the average weighted length of the post-initial phrase, fp the frequency of the particular enclitic in post-initial position, fn the frequency of the particular enclitic in non-post-initial position

<i>sě</i>				<i>mi</i>				<i>tě</i>			
ALi	fp	fn	pp	ALi	fp	fn	pp	ALi	fp	fn	pp
1	229	112	0.67	1	189	8	0.96	1	91	13	0.88
2	139	119	0.54	2	46	6	0.88	2	58	9	0.87
3	16	85	0.16	3	3	4	0.43	3	4	10	0.29
4	3	53	0.05	5.14	1	6	0.14	4.33	3	6	0.33
6.14	1	21	0.05								

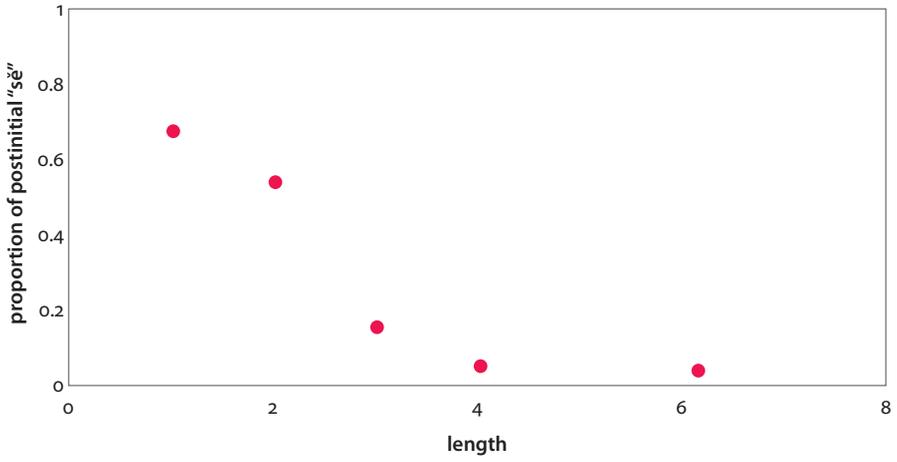


Figure 4. Proportions of post-initial positions of enclitic *sě*. The length of the initial phrase is measured by the number of words

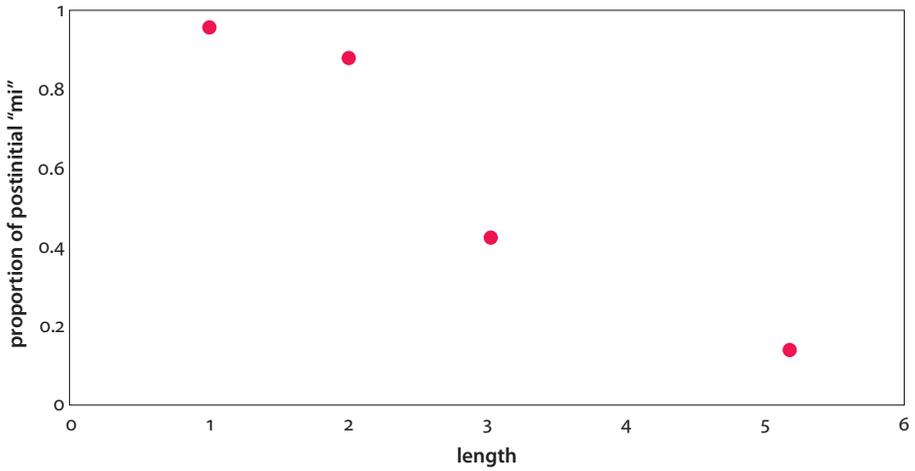


Figure 5. Proportions of post-initial positions of enclitic *mi*. The length of the initial phrase is measured by the number of words

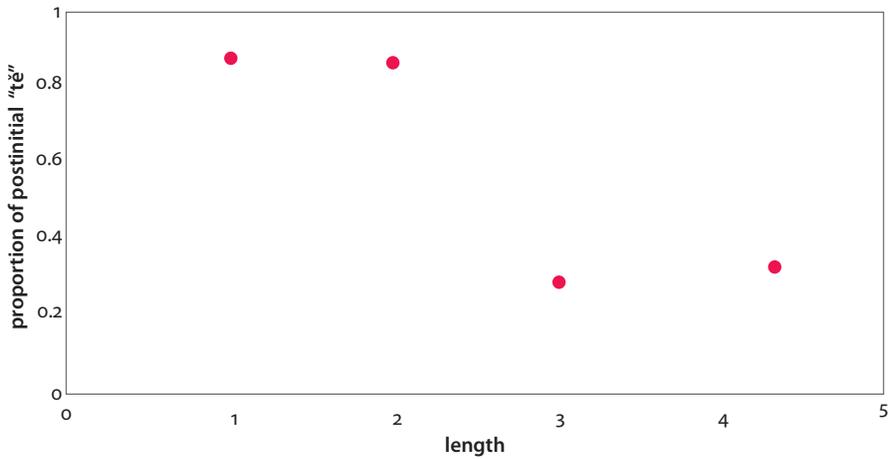


Figure 6. Proportions of post-initial positions of enclitic *tě*. The length of the initial phrase is measured by the number of words

6. Conclusion

In this study, the impact of the length of the initial phrase on the position of four enclitics in Old Czech was analyzed. The hypothesis predicting the negative correlation between the length of this phrase and the proportions of enclitics in the post-initial position was not rejected. Moreover, the study reveals that the general tendency, i.e., the longer the phrase, the lower the proportion of enclitics in post-initial position, manifests differently in the case of *sě*, on the one hand, than in the case of *mi* and *tě*, on the other. We suppose that this difference can be caused by the following factors. First, both *mi* and *tě* represent ‘pure’ pronouns whose function is to denote an object, while *sě* is grammatically multifunctional. Specifically, *sě* occurs as (1) a pronoun, (2) a grammatical morpheme, which performs various grammatical functions (primarily deagentization and intransitivization), (3) a discrete part of lexical items (reflexiva tantum). This language unit shares functions of a grammatical morpheme as well as of a pronoun and, thus, ‘lies’ on the border of these functions. Second, the results can be influenced by a word order position of a syntactically superordinate verb in a clause. Our preliminary findings (Kosek et al. 2018b) show that a fine-grained classification of word order can shed light on some characteristics of the observed phenomena. To sum up, despite all differences, there seems to be a clear tendency in accordance with the hypothesis.

Funding

This study was supported by the project *Development of the Czech pronominal (en)clitics* (GAČR GA17-02545S).

© 2021. John Benjamins Publishing Company

All rights reserved

References

- Čavar, Damir & Chris Wilder. 1999. "Clitic Third" in Croatian. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 429–467. Berlin: de Gruyter.
<https://doi.org/10.1515/9783110804010.429>
- Ertl, Václav. 1924. Příspěvek k pravidlu o postavení příklonek. *Naše řeč* 8(9), 257–268; 8(10), 293–309.
- Kosek, Pavel, Olga Navrátilová, Radek Čech & Ján Mačutek. 2018a. Word order of reflexive 'se' in finite verb phrases in the first edition of the Old Czech Bible translation (Part 1). *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 135. 177–188.
<https://doi.org/10.4467/20834624SL.18.017.8853>
- Kosek, Pavel, Olga Navrátilová & Radek Čech. 2018b. Slovosled prominálních enklitik mi, si, ti, ho, mu v Bibli kralické. Paper presented at Diachronní setkání v Řevnicích, Řevnice, October 31–November 2.
- Kosek, Pavel, Olga Navrátilová & Radek Čech. 2019. *The development of the Czech pronominal (en)clitics*. <http://www.cechradek.cz/enclitics/home.html>
- Kyas, Vladimír. 1997. *Česká Bible v dějinách národního písemnictví*. Prague: Vyšehrad.
- Kyas, Vladimír (ed.). 1981. *Staročeská bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu bible ze 14. století. I. Evangelia*. Prague: Academia.
- Kyas, Vladimír (ed.). 1985. *Staročeská bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu bible ze 14. století s částmi Bible litoměřicko-třeboňské. II. Epištoly. Skutky apoštolů. Apokalypsa*. Prague: Academia.
- Kyas, Vladimír (ed.). 1988. *Staročeská bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu bible ze 14. století. III. Genesis–Esdráš*. Prague: Academia.
- Kyas, Vladimír, Věra Kyasová & Jaroslava Pečirková (eds.). 1996. *Staročeská bible drážďanská a olomoucká: kritické vydání nejstaršího českého překladu bible ze 14. století. IV. Tobiaš–Sira-chovec*. Paderborn: Schöningh.
- Pečirková, Jaroslava, Hana Sobalíková, Markéta Pytlíková, Milada Homolková, Vladimír Kyas & Věra Kyasová (eds.). 2009. *Staročeská Bible drážďanská a olomoucká s částmi Proroků žmberských a Bible litoměřicko-třeboňské. V/1 Izaiáš–Daniel, V/2 Ozeáš–2. kniha Makabejská*. Prague: Academia.
- Radanović-Kocić, Vesna. 1996. The placement of Serbo-Croatian clitics: A prosodic approach. In Aaron Halpern & Arnold Zwicky (eds.), *Approaching second: Second position clitics and related phenomena*, 429–445. Stanford, CA: CSLI Publications.
- Šlosar, Dušan. 1967. Poloha enklitik jako kritérium k hodnocení staročeské interpunkce. *Listy filologické* 91(3). 251–258.
- Toman, Jindřich. 2000. Prosodické spekulace o klitikách v nekanonických pozicích. In Zdena Hladká & Petr Karlík (eds.), *Čeština – univerzálie a specifika* 2, 161–166. Brno: Masarykova univerzita.
- Trávníček, František. 1956. *Historická mluvnice česká 3. Skladba*. Prague: SPN.
- Uhlířová, Ludmila, Petr Kosta & Ludmila Veselovská. 2017. Klitikon. In Petr Karlík, Marek Nekula & Jana Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. <https://www.czechency.org/slovník/KLITIKON> (30 December, 2018)
- Vintr, Josef. 2008. *Bible (staroslověnský překlad, české překlady)*. In Luboš Merhaut (ed.), *Lexikon české literatury, 4/II U–Ž, Dodatky A–Ř*, 1882–1887. Prague: Academia.
- Zwicky, Arnold. M. 1994. What is a clitic? In Joel A. Nevis, Brian D. Joseph, Dieter Wanner & Arnold M. Zwicky (eds.), *Clitics. A comprehensive bibliography 1892–1991*, 12–20. Amsterdam: Benjamins. <https://doi.org/10.1075/lisl.22>