Why are there variations of the word order positions of enclitics in Old Czech?

Radek Čech, Pavel Kosek, Olga Navrátilová

This paper was supported by the project Development of the Czech pronominal (en)clitics (GAČR GA17-02545S)
Research background

Kosek & Navrátilová

- long-standing research of Old and Middle Czech
- “classical” philological approach + frequency characteristics
- manually annotated corpora
- detailed description of various features of Old and Middle Czech
- a need to find more general mechanisms
- an explanation
Research background

Kosek & Navrátilová + Čech

● theoretical background of QL
● the least effort principle & synergetic linguistics
● the Menzerath-Altmann law
Enclitics

- language units that are phonetically dependent on preceding word
- there must be no syntactic relationship between enclitic and preceding word
- form
  - discrete morpheme
  - auxiliary verb
  - pronoun

V zahradě se starý strom náhle skácel
In the garden, an old tree suddenly fell
Enclitics (in verb finite phrase) in Contemporary and Old Czech

- **Contemporary Czech**
  - relatively stable position = after the first phrase of clause (from the first half of 20. century)

- **Old (and Middle) Czech**
  - word order variation
    - after the first word / phrase of clause
    - in the middle / at the end of clause - connected to verb (in almost all cases)
    - etc.

- **research questions**
  - are there any regularities in a word order distribution of enclitics in Old Czech?
  - how are particular word order positions connected to other language properties?
Language material

- chosen books of the oldest Czech Bible translation Bible olomoucká (1417)
  - Old Testament (Gen, Isa, Job, Ecc)
  - New Testament (Mt, Lk, Act, Rev)

- transcription and manual annotation
Description, generalization and stochastic rules

- R1: if an enclitic appears in a clause, use it after the initial phrase / word of the clause
- R2: if rule (R1) is not applied, use the enclitic in a postposition of a verb
- R3: if rule (R2) is not applied, use the enclitic in a pre-position of a verb
Description, generalization and stochastic rules

- R1: if an enclitic appears in a clause, use it after the initial phrase / word of the clause
- R2: if rule (R1) is not applied, use the enclitic in a postposition of a verb
- R3: if rule (R2) is not applied, use the enclitic in a pre-position of a verb

QUESTION

- are there any conditions which can influence application of particular rules?
Length of the initial phrase of a clause

- the enclitic almost never occur after a pause
- the longer the first phrase of the clause, the higher probability that the pause is realized
Length of the initial phrase of a clause I

- the enclitic almost never occur after a pause
- the longer the first phrase of the clause, the higher probability that the pause is realized

HYPOTHESIS

The longer the first phrase of the clause, the lower probability of the occurrence of the enclitic after this phrase
Methodology

● phrases are determined syntactically
  ○ in Contemporary Czech syntactic phrase usually corresponds with an intonational phrase (in phonetics)
● the length of the phrase measured in a number of
  ○ letters
  ○ words
● clauses containing enclitics in any position used for the analysis
● for phrases of length \( L = x, x = \{1, 2, 3, \ldots n\} \), a proportion of phrases with enclitics in position after the first phrase is measured
Methodology

- enclitic “sě”
  - the most frequent enclitic in our corpus
- 761 clauses from *Bible Olomoucká* (Gen, Isa, Job, Ecc, Mt, Lk, Act, Rev)
- data are pooled - for each group at least 10 occurrences
- weighted length (a weight is a frequency)
Results - letters

The graph illustrates the proportion of postinitial "se" as a function of length. The y-axis represents the proportion of postinitial "se," while the x-axis represents the length. As the length increases, the proportion of postinitial "se" decreases.
Results - words

![Graph showing the proportion of postinitial "se" as a function of length.](image)
Length of the initial phrase of a clause II

- properties of initial phrase
  - in clauses with enclitics after this phrase (LiP and LiP_se)
  - in clauses with enclitics which do not occur after this phrase (LiN)
  - in clauses without enclitics (LiW)
Methodology

- enclitic “sě”
  - the most frequent enclitic in our corpus
- books *Matthew* and *Job* from *Bible Olomoucká*
- length of the initial phrase measured
  - in letters
  - in words
- differences tested by Mann-Whitney-Wilcoxon test
Results - letters (Matthew)
Results - letters (Job)
Results - words (Matthew)
Results - words (Job)

- LiP: 1.29
- LiP_se: 2.29
- LiN: 2.22
- LiW: 1.14
Results

- clauses without the enclitic significantly shorter than clauses with the enclitic in any position (for both letters and words)
- the enclitic significantly lengthens the initial phrase
Enclitics and clause properties

● an occurrence of the enclitic and properties of whole clause
● mean length of the clause
  ○ measured in a number of phrases directly dependent on a predicate

● mean length of phrases of the clause
Methodology

- enclitic “sě”
  - the most frequent enclitic in our corpus
- books *Matthew* and *Job* from *Bible Olomoucká*
- 60 occurrences
Results - mean L of clauses (*Matthew*)

- all differences highly significant ($p < 0.001$)
Results - mean L of clauses (Job)

- all differences highly significant ($p < 0.001$)
Results - mean L of phrases (Matthew)

![Bar chart showing mean length of phrases for LPP, LPN, and LPW with values 2.57, 2.23, and 2.84 respectively.]
Results - mean L of phrases (Job)

![Bar chart showing mean length for LPP, LPN, and LPW]
Results - mean L of clauses and phrases (Matthew)
Results - mean L of clauses and phrases (Job)
Conclusions

- systematic relationships between the length of the initial phrase and the word order position of enclitics
- clauses with enclitics have longer initial phrase than clauses without enclitics
- clauses with enclitics are significantly longer than clauses without enclitics
- clauses with enclitics which do not occur after initial phrase are significantly longer than clauses with enclitics after the initial phrase
- longer clauses have shorter phrases (cf. the Menzerath-Altmann law)
Thank you!