

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

in garden LOC.M.SG **REFL** ACC *old* NOM.M.SG *tree* NOM.M.SG *suddenly fall* PART.PRET.ACT.M.SG

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

a a foueratawre nase prugil apo
 tot geho arone tasto se poayna
Do gmagisse acenij S. mathege



Blychj naroceny genu
 kyfsta fyna dan dowi fy
 na abrahamona. U bna
 ham wrodil yna fia. U pu
 ak wrodil iakuba a yafub wrodil yu
 du a bnaty geho. Wyda pak wrodil
 ybarte a tarumna. zec aceny gneto
 gme byfo thamar. U barte pak v
 wrodil arona. Eron pak wrodil au
 ma. U eron pak wrodil amyna dabi
 Amyna dab wrodil na arona. U la
 ron pak wrodil falumona. U falumon
 pak wrodil booez. zec aceny gneto g
 me bieffe maab. Booez pak wrodil ode
 tha. zec aceny gneto gme byfo nich
 U o beth pak wrodil yeffe. U yeffe pak
 wrodil firale dauida. Dauid pak firal
 wrodil falumona. zec aceny gneto me
 ky by falumona fudy firale dauidona
 genuwto gme wryaf. Galumon pak
 wrodil robdama. U robdam pak wro
 dil abryama. U abas pak wrodil aza.
 U azu pak wrodil yocapham. U yocaph
 ar pak wrodil yocama. Yocam pak wro
 dil oza. Oza pak wrodil yocathana
 Do arban wrodil achana. U achar pak v
 wrodil ezechiaffe. Ezechiaff pak wrodil
 manaffon. U manaffes pak wrodil amo
 na. amon pak wrodil yocuffie. U yocuf
 pak wrodil yocuffie a bnaty geho
 U to bit rodo aafi. U byr fit byr wydne
 v waceny w babylonie. U po babylon
 fien waceny fi dyaufie u: dee wrauli

do fivereme. U fion yafu pak wrodil
 fulanick. Galanek pak wrodil wroba
 beke. U wroba beke pak wrodil abryudi.
 U bryud pak wrodil achyma. achachi
 pak wrodil arona. U aron pak wrodil fa
 docha. U fadoch pak wrodil achyma.
 U achym pak wrodil eluda. U elud pak
 wrodil ekeazam. Ekeazar pak wrodil
 mathana. U mathan pak wrodil yafu
 ba. U afub pak wrodil yocepha mme
 fivare marie. U yocuffie gefit narodil
 genuf genuwto diegy firfius. U firf
 fany narodonec or abrahamona ar do
 dauida narodow arnadare. U or da
 uida ar do giene babylon ftecho nar
 dow arnadare a ortiene babylon
 ftecho ar do naftecho iculit fit nar
 dow arnadare. U naroceny pak na
 ftecho iculit fit takti fit gefit ftefo
 U dyr bieffe ymanzef fivo danamar
 fia naftecho genuf fita mana yocepfo
 wri. U wri waceny fte ftecho celefitye
 na ftecha by re gefit yocafu wrywo
 te magna. a to acen byfo yocafu w
 wryw fteche ducha fivarecho bieffe. U eb
 dy yocepf mmi genu re bieffe muz
 fprandim nechac gnet ftecho donu
 wryw fteha a zgiw wryw yocafu yna eche
 fte gnet wryw ofte by fteha. U fi dneto
 myf fteff. U eb fi angel boe zgiw fte
 genu wryw a fia takti. U yocephe
 fynu dauidow romer wodi dauidona
 u eb or fte prugien man fiverem y led
 to acen fte gefit wryw narodilo. to gefit
 fteche ducha fivarecho. U eb wrodil fyna
 a wrodil gefit genuf genuf gme roat
 fte fteche. U eb on fte fte fte fte fte fte
 che w gich. a to fte w fte fte fte w oco ady
 fte w aply fo fte fte poto fia yz a yafte
 w fte dnet fte aply fte gnet yz a u a fia
 takti. U y pama pome wryw ome
 fte w y porodi fyna genuwto gme
 bnd emanuel. to fte w fte fia dnet
 fte am. U eb dy yocepf w fte w fte fte
 wryw fte fia genuf bieffe w fte fia
 angel boe. U wri fte wryw a nopo
 na gne celefitye. ar ona wodi fte
 bo fyna pwoocentcho y ptece. die
 che genuf genuf. **U in by wrodil.**

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

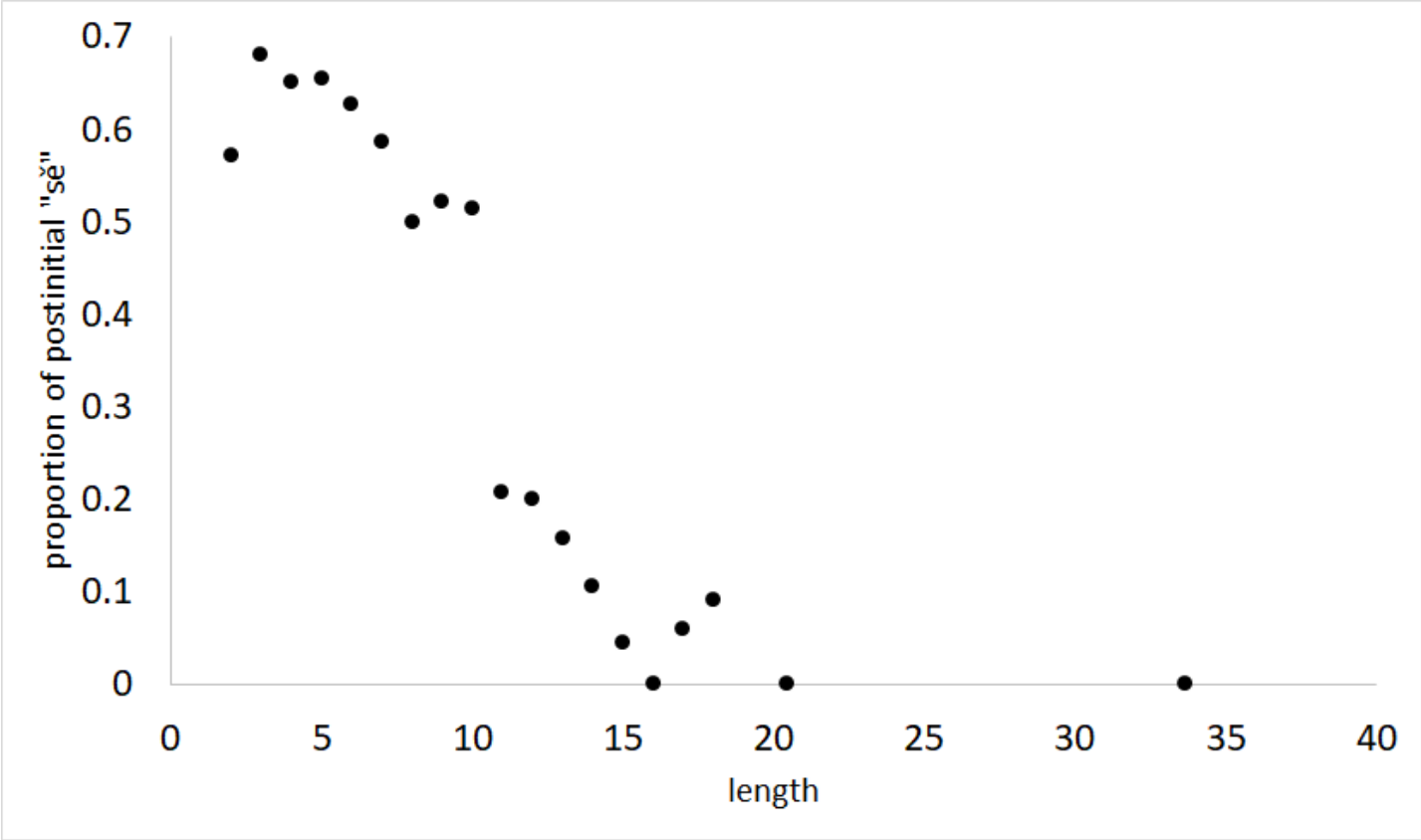
Methodology

- 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 $L = x$, $x = \{1, 2, 3, \dots n\}$,
a proportion of phrases with enclitics in position after the first phrase is measured

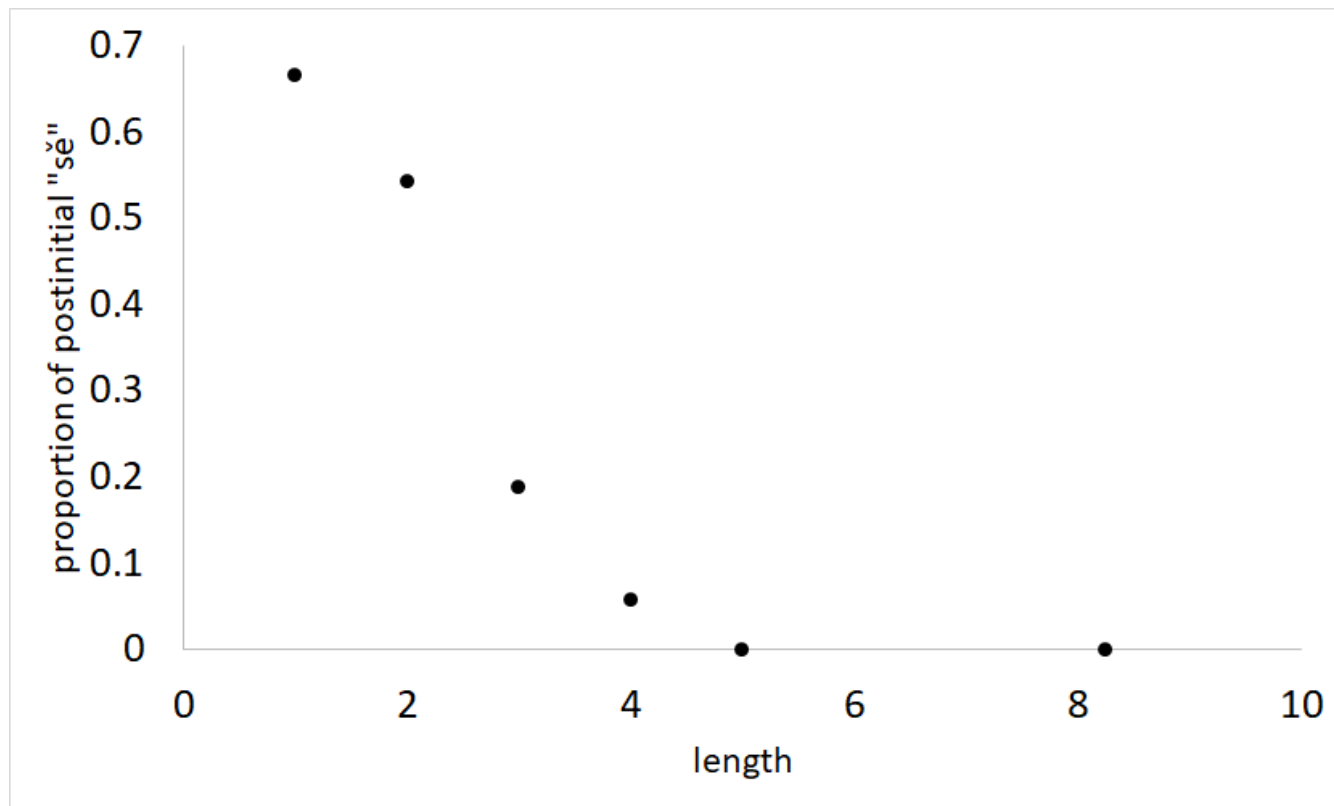
Methodology

- 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



Results - words



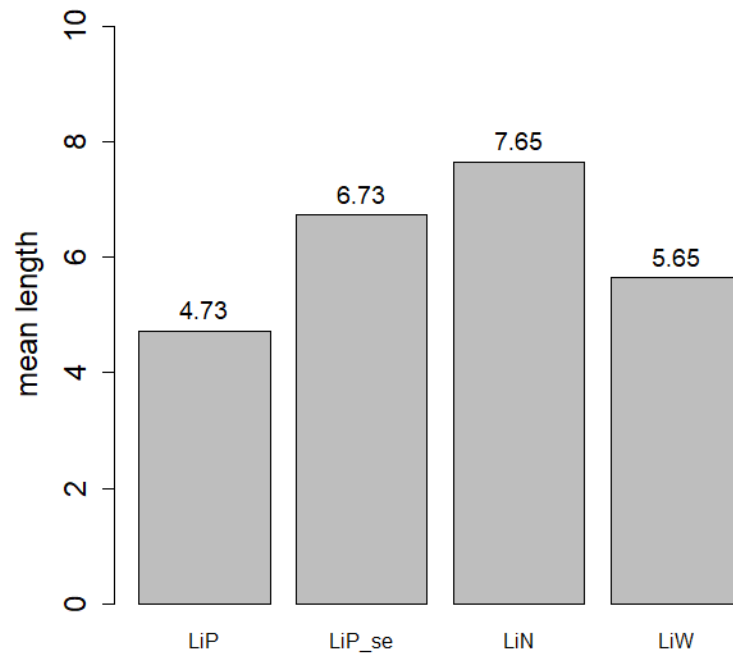
Length of the initial phrase of a clause II

- 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)

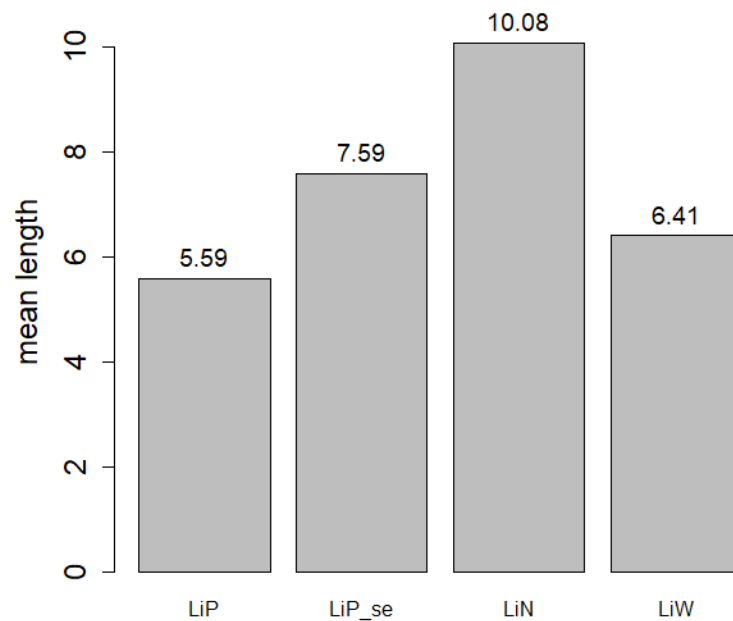
Methodology

- 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

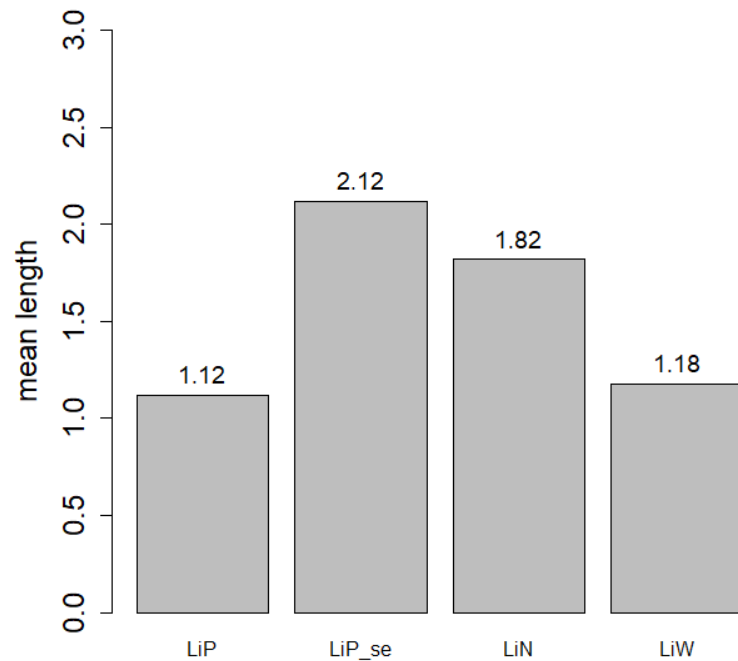
Results - letters (Matthew)



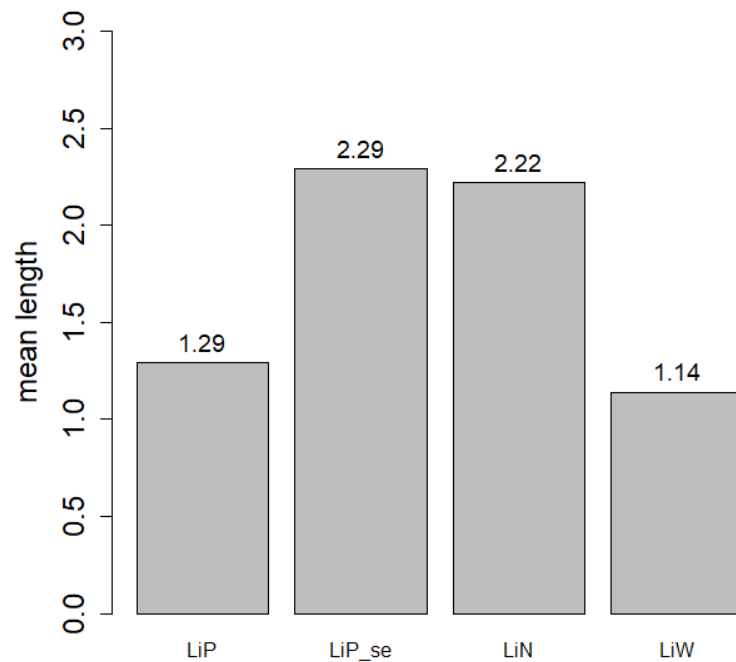
Results - letters (Job)



Results - words (Matthew)



Results - words (Job)

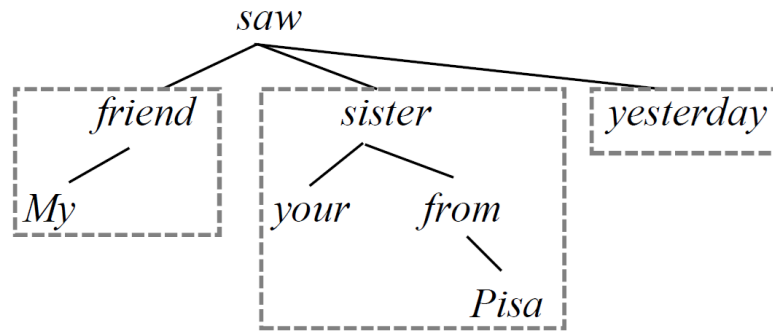


Results

- 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

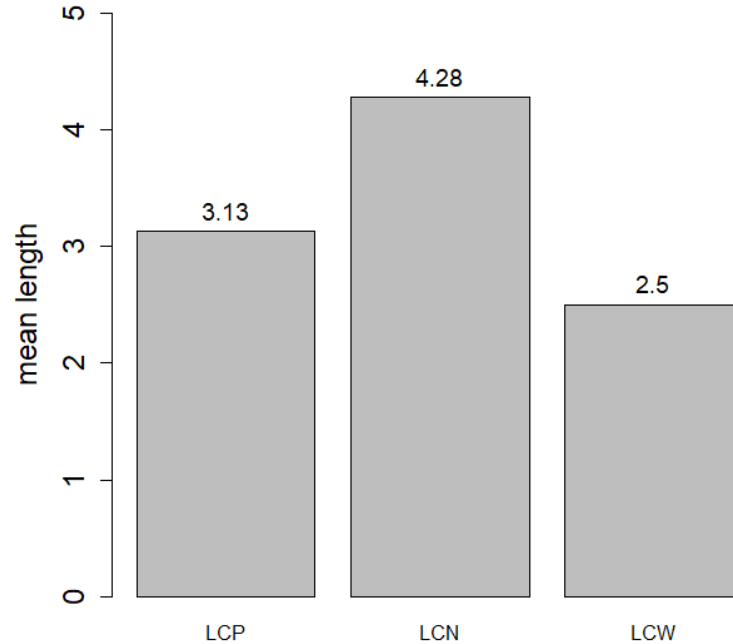


- mean length of phrases of the clause

Methodology

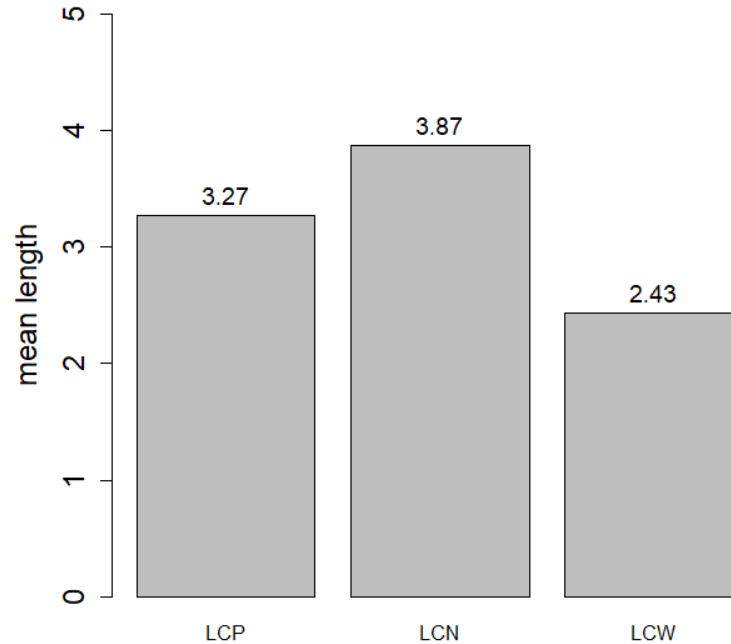
- enclitic “sě”
 - the most frequent enclitic in our corpus
- books *Matthew* and *Job* from *Bible Olomoucká*
- 60 occurrences

Results - mean L of clauses (*Matthew*)



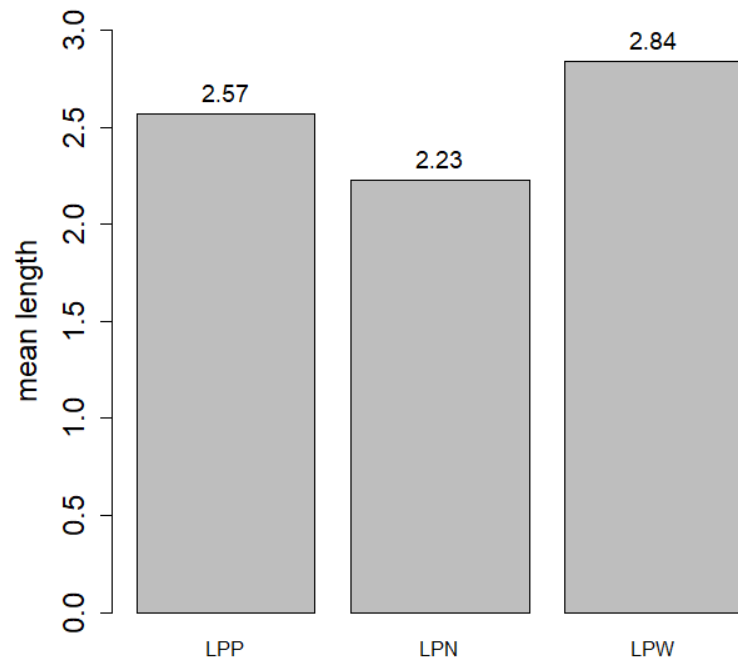
- all differences highly significant ($p < 0.001$)

Results - mean L of clauses (*Job*)

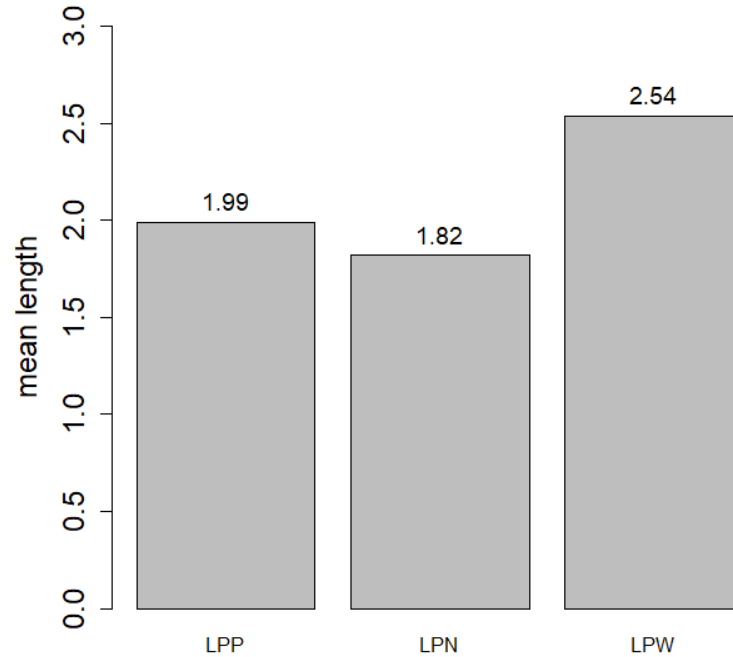


- all differences highly significant ($p < 0.001$)

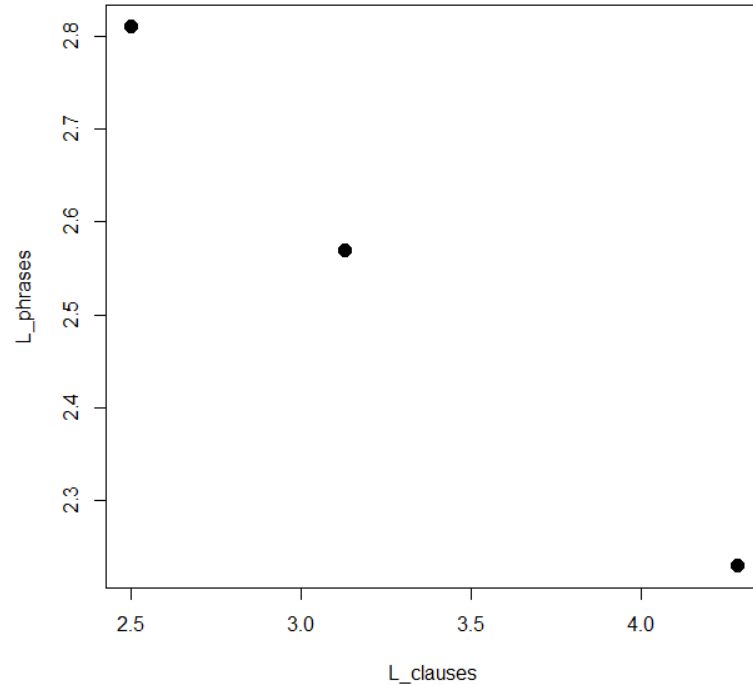
Results - mean L of phrases (*Matthew*)



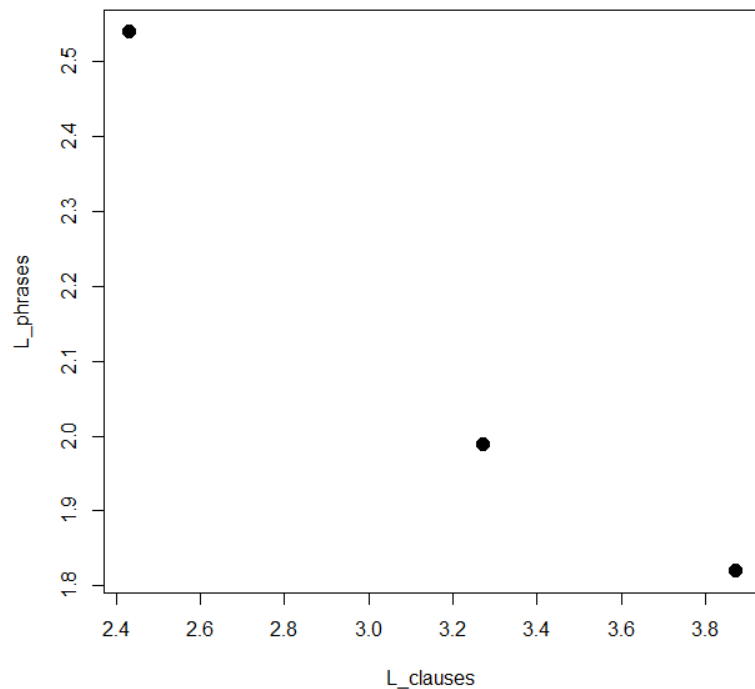
Results - mean L of phrases (*Job*)



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!