Gemination and clitic order in Czech: historical development and theoretical insights

Introduction. In standard Present-Day Czech (PDCz), clitics follow a rigid order, illustrated schematically in (1), such that only (2a) represents a grammatical order of clitics in the clause (the preterite auxiliary is glossed as an exponent of person & number). Any other permutation, e.g., (2b)–(2c), or splitting the cluster with an intervening element (2d), is not allowed. In 2SG, a non-syllabic form of the preterite auxiliary, *s*, can be used instead of syllabic *jsi* and is even preferred by speakers of some dialects (ČJA: 594f). Importantly, this form seems to occupy the same position in the cluster (2e).

(1)   AUX(ILIARY)P(RETERITE) » REFL(EXIVE) » DAT(IVE) » ACC(USATIVE)   (PMČ: 649)

(2)   a.   *Poslal/i   **jsem/jsi/jsme/jste   mu   to**   určitě   včera.*
             sent.MSG/PL 1SG/2SG/1PL/2PL him.DAT it.ACC definitely yesterday
             'I/you/we have sent it to him definitely yesterday.'

        b.   **Poslal/i to mu jsem/jsi/jsme/jste** *určitě včera.*
        c.   **Poslal/i mu to jsem/jsi/jsme/jste** *určitě včera.*
        d.   **Poslal/i jsem/jsi/jsme/jste** *určitě* **mu to** *včera.*
        e.   *Poslal **s mu to** určitě včera.*

Furthermore, when a clause contains the auxiliary clitic and a reflexive clitic *se*.REFL.ACC (similarly with *si*.REFL.DAT), the grammatical sequence usually follows from (1), see (3a). In 2SG, however, a reversed order of clitics is necessary for the sequence to be grammatical, cf. (3b)–(3c); but only the non-syllabic form is permitted (3d), whereas the syllabic *jsi* is not (3e).

(3)   a.   *Ztrapnil/i   **jsem/jsme/jste se.***
             embarrased.MSG/PL 1SG/1PL/2PL REFL.ACC
             'I/we/you embarrassed myself/ourselves/yourselves.'

        b.   **Ztrapnil   jsi   se.**
             embarrased.MSG 2SG REFL.ACC
             'You embarrassed yourself.'

        c.   **Ztrapnil s se.**
        d.   *Ztrapnil **se s**.*
        e.   **Ztrapnil se jsi.**

Even though the standard language disallows it, speakers of some dialects exhibit the sequences in (3b)–(3c) as predicted by (1) (ČJA: 594f, NOVÁKOVÁ 2018). Moreover, as Gebauer (1929: §134) shows, these sequences were fairly common in earlier stages of Czech. Clearly, the standard sequence in (3d) is a historical novelty that was promoted in the standard language.

Aim. While these general observations were made earlier (see especially Sedláček 1994), to my knowledge, no detailed account of the phenomenon exists. For instance, Franks & King (2000: 346f) capture its structural aspect, however, the reason behind the position swap in 2SG remains unclear. Sedláček (1994) builds on previous literature and claims that the change from *s se* to *se s* is due to an articulatory challenge posed by the geminate in the original sequence *s se*. I follow up on this claim and aim to support it with empirical data. For this purpose, I investigate the development of three different morphosyntactic contexts known to contain geminated *s* in the Old Czech era (ca. 14c.–15c.), which were later resolved in PDCz: 1. prefix (*ssaditi → sesadit* 'dethrone'), 2. preposition (*s synem → se synem* 'with a son'), and 3. clitics (*s se → se s*). The aim is to determine whether the resolution of geminated *s*—whether through vowel epenthesis in prefixes and prepositions or positional swap in clitics—occurred around the same time in the language's history, suggesting that both processes might have shared a common trigger.

Results. The queries to the Czech diachronic corpus, DIAKORP, yielded a total of 1,256 instances across all three contexts in written texts, dating from the 14c. to the first half of the 20c. The ratio of geminated (ss) and degeminated (ses) forms in the figure below shows that geminated *s* predominated in the Old and Middle Czech periods, up until the end of the 18c. In the 19c., a rapid decrease in gemination occurred across all observed contexts. The log-linear model analysis suggests that vowel adjacency was an important factor during the 19c., similarly to contemporary Polish (Pająk 2009), with vowel-adjacent

geminates (e.g., ***ssaditi***) resisting the change for several decades longer than non-vowel-adjacent geminates (e.g., ***sstúpiti*** 'descend'). The effect of the morphosyntactic category (clitic–prefix–preposition), on the other hand, was insignificant, despite residual geminates persisting into the first half of the 20c. (most of which could be explained as orthographic exceptions). This supports previous claims in the literature that the partial reorganization of the Czech clitic cluster resulted from the resolution of geminated *s*, with different solutions emerging at other levels of grammar.
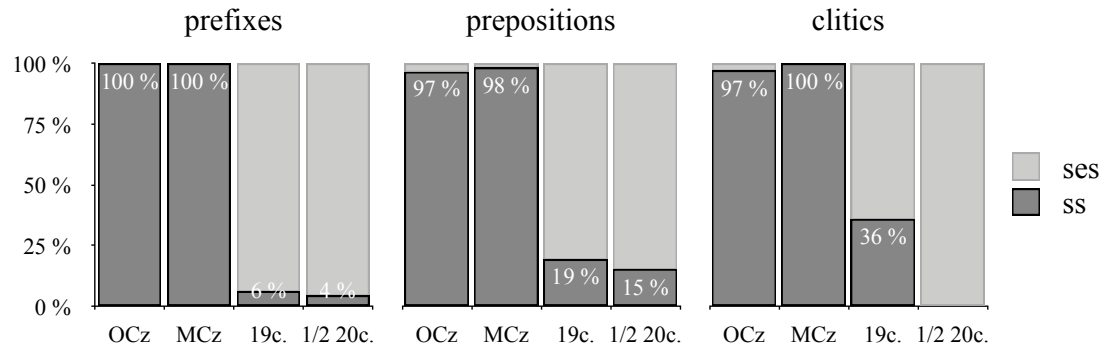
Analysis. In a tentative analysis of the observed phenomena, I employ an Optimality-Theoretic approach (Prince & Smolensky 2004). The diachronic change is captured directly by the rise of constraints against geminates during the 19c. To explain the difference between prefixes and prepositions on one side (vowel epenthesis) and clitics on the other (position swap), I adhere to Stratal OT (Kiparsky 2015) and assume that at the lower level of grammar (presumably the word stratum), vowel epenthesis is favored due to the lower ranking of the DEP(V) constraint. In contrast, at the higher level (presumably the phrase stratum), DEP(V) dominates alignment constraint(s) that would otherwise force the auxiliary to precede the reflexive (see the LEH constraint in Franks & King 2000: 341f). Finally, the high-ranked REALMOR constraint (e.g., Pająk 2009) prevents elision of an entire morpheme as another means of resolving geminated *s*. The rankings in (4) summarize the proposed hierarchy of constraints at different grammatical strata. Additionally, I sketch possible analyses for future research. For instance, the different behavior of word-medial geminates (*rozzlobit* 'get angry'), which resist vowel epenthesis, in contrast to the word-initial geminates discussed here, could be addressed by splitting the *GEM constraints into two variants. I also tentatively suggest a more general constraint derived from the Sonority Sequencing Principle (e.g., Clements 1990).

(4) a. REALMOR » LEH » {*GEM} » DEP(V) » {*GEM}   (lower level)
     b. REALMOR » DEP(V) » {*GEM} » LEH » {*GEM}   (higher level)

References. ČJA – *Český jazykový atlas* [Czech Linguistic Atlas]. Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i, Praha, 2014. CLEMENTS, G. N. (1990). The role of the sonority cycle in core syllabification. In Kingston, J., & Beckman, M. E. (eds.), *Papers in laboratory phonology 1*, pp. 283–333, Cambridge: Cambridge University Press. DIAKORP – KUČERA, K., *et al. Diakorp: Diachronní korpus* [Diakorp: Diachronic Corpus], v6, Ústav Českého národního korpusu FF UK, Praha, 2015. FRANKS, S., & KING, T. H. (2000). *A Handbook of Slavic Clitics*, New York: Oxford University Press. GEBAUER, J. (1929). *Historická mluvnice jazyka českého, Díl IV, Skladba* [Historical Grammar of the Czech Language, Part IV, Syntax], Praha: ČAVU. KIPARSKY, P. (2015). Stratal OT: A synopsis and FAQs. In Hsiao, Y. E., & Wee, L.-H. (eds.), *Capturing Phonological Shades Within and Across Languages*, pp. 2–44, Newcastle upon Tyne: CSP. NOVÁKOVÁ, B. (2018). *Řazení klitiky -s v současné češtině* [Placement of Clitic -s in Contemporary Czech], M.A. Thesis, Brno: Masaryk University. PAJĄK, B. (2009). Contextual Constraints on Geminates: The Case of Polish, *BLS* 35(1), pp. 269–280. PMČ – Grepl, M., *et al.* (2012). *Příruční mluvnice češtiny* [Reference Grammar of Czech], Praha: NLN. PRINCE, A., & SMOLENSKY, P. (2004). *Optimality theory: Constraint Interaction in Generative Grammar*, Malden: Blackwell Publishing. SEDLÁČEK, M. (1994). Ty jsi se (učil), ty jsi si (pamatoval)? [You learned, you remembered?] *Naše řeč*, 76(1), pp. 27–43.