

VARIATION IN THE CASE FORMS OF THE INDEFINITE PRONOUN *KEEGI* 'SOMEONE': A COMPARATIVE CORPUS STUDY OF WRITTEN AND SPOKEN ESTONIAN

Annika Kängsepp

University of Tartu

Keywords: corpus linguistics, language variation, indefinite pronouns, morphology, Estonian

Variation in the case forms of the Estonian indefinite pronouns *keegi* 'someone', *miski* 'something', *kumbki* 'either', and *ükski* 'none' has been observed in written language (Rull 1917) and dialects (Saareste 1955) for over a century. This variation concerns the placement of the *-gi/-ki*, which is typically regarded as an emphatic clitic in other contexts. Within indefinite pronouns, however, it can occur after the case ending (e.g., *kelle-le=gi* someone-ALL=CLIT 'to someone'), before the case ending (e.g., *kellegi-le* someone.INDF-ALL), between two case endings (e.g., *kelle-le-gi-le* someone-ALL-INDF-ALL) or before and after the case ending (e.g., *kellegi-le=gi* someone.INDF-ALL=CLIT) (Saareste 1923, 1936). In indefinite pronouns, *-gi/-ki* is perceived more as part of the stem. The variation has a strong dialectal background, as forms where *-gi/-ki* is placed after the case ending have historically been common only in Southern and Northeast Estonia (Saareste 1955). Nevertheless, in standard Estonian, the normative placement occurs only after the case ending.

As this variation has been systematically studied very little and only in written contemporary Estonian (Pant 2018, 2020), this presentation aims to provide an overview of its extent and to identify the factors influencing the variation in the case forms of the pronoun *keegi* through a comparative analysis of written and spoken Estonian. The analysis of written language is based on data from the Estonian National Corpus (2.4 billion words; Koppel, Kallas 2022). For the analysis of spoken language, data were drawn from two corpora: the Estonian Public Broadcasting's Radio Corpus (109 million words; Lippus et al., 2023a), and the Estonian Podcast Corpus (85 million words; Lippus et al., 2023b). To examine the variation, the proportion of occurrences was calculated, while statistical analysis was applied to identify the factors influencing the placement of *-gi/-ki*.

The findings reveal that in written Estonian, *-gi/-ki* is predominantly placed after the case ending, accounting for 78.6%, while forms where *-gi/-ki* precedes the case ending constitute 20.6%. Instances where *-gi/-ki* appears between case endings or both before and after it represent the remaining 0.8%. In spoken Estonian, the distribution is more varied, with *-gi/-ki* appearing after the case ending in 54.2% of occurrences, before the case ending in 43.4%, and in other positions in 2.4%. Univariate analysis of written Estonian identified genre, the occurrence of the pronoun as an attribute, and the function of the pronoun in a clause as significant factors influencing variation. In spoken Estonian, the corpus, the position of the pronoun in the clause, the gender of the speaker, case, and speech rate were found to significantly affect the placement of *-gi/-ki*. Multivariate analysis further indicated that genre, and whether the text had been edited, was the strongest factor influencing written language variation. In spoken language, speech rate emerged as the most influential factor, with faster speech favouring the placement of *-gi/-ki* before or between two case endings. Additionally, male speakers were more likely to produce forms where *-gi/-ki* precedes or occurs between case endings (Kängsepp 2024, 2025).

This presentation argues that corpus-based analysis provides valuable insight into grammatical variation across registers, emphasizing how using both corpora offers a more comprehensive view by incorporating linguistic, sociolinguistic, and prosodic factors.

This work was supported by the National Program: Estonian Language and Culture in the Digital Age project “Morphosyntactic variation in Estonian” (EKKD-TA2).

References

- Koppel, Kristina, Kallas, Jelena. 2022. Eesti keele ühendkorpuste sari 2013–2021: mahukaim eesti keele kogu [Estonian National Corpus 2013–2021: The largest collection of Estonian language data]. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics*. 18, 207–228. <https://doi.org/10.5128/ERYa18.12>
- Kängsepp, Annika. 2024. Käänevormide varieerumine ja seda mõjutavad tegurid kirjalikus keeles indefiniitpronoomeni keegi näitel [The variation in the case forms of the indefinite pronoun keegi ‘someone’ in written Estonian]. *Keel ja Kirjandus*. 11, 1016–1037. <https://doi.org/10.54013/kk803a3>
- Kängsepp, Annika. 2025. Indefiniitpronoomenite keegi ja miski käänevormide varieerumine suulises keeles [Variation in the case forms of the indefinite pronouns keegi ‘someone’ and miski ‘something’ in spoken Estonian]. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics*. 21, 121–139. <https://doi.org/10.5128/ERYa21.07>