

Variation in the Case Forms of Indefinite Pronouns: A Multi-Method Approach Combining Corpus Studies and Eye-Tracking

Annika Kängsepp

Variation in the case forms of the Estonian indefinite pronouns *keegi* ‘someone’, *miski* ‘something’, *kumbki* ‘either’, and *ükski* ‘none’ has been documented for over a century, both in written language (Rull 1917) and dialects (Saareste 1955). This variation concerns the placement of the *-gi/-ki*, which 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 both before and after the case ending (e.g., *kellegi-le=gi* someone.INDF-ALL=CLIT) (Saareste 1923, 1936). The variation has a strong dialectal background: forms where *-gi/-ki* is placed after the case ending have historically been prevalent only in Southern and Northeast Estonia (Saareste 1955). Despite this, in standard Estonian, the normative placement of *-gi/-ki* occurs only after the case ending.

Given that this variation has been systematically studied only in written Estonian (Pant 2018, 2020), this presentation explores the variation using written and spoken data, with a combination of corpus studies and eye-tracking. The aim of corpus studies was to provide an overview of the extent of the variation and describe the factors influencing the choice between the normative and non-normative forms. Since written language limits the factors that can be examined, studying spoken language was necessary to understand the impact of sociolinguistic factors and prosody. While the corpus studies provide insights into the production of forms, eye-tracking offers a complementary view into real-time processing of varying forms by tracking eye movements such as fixations and regressions during reading (e.g., Holmqvist et al., 2011). Taking a broader perspective, the eye-tracking experiment explores how normativity and frequency affect the reading of morphologically varied forms.

The corpus study of written language utilized data from the Estonian National Corpus 2021 (2.4 billion words; Koppel & Kallas 2022). For the corpus study of spoken language, data was drawn from 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). The experimental stimuli were based on natural examples from the written corpora and further manipulated to obtain comparable context length and structure.

The results from the corpus studies indicate that in written Estonian, *-gi/-ki* is predominantly placed after the case ending (78.6%), whereas in spoken Estonian, it occurs at almost equal frequency after (54.2%) and before (43.3%) the case ending. In written language genre, and whether the text had been edited, was the strongest factor influencing variation. In spoken language, speech rate emerged as the most influential factor, with faster speech favoring the placement of *-gi/-ki* before or between two case endings. (Kängsepp 2024, 2025) Regarding processing, the hypothesis for the ongoing eye-tracking study posits that there are significant differences between how normative and non-normative forms are read, especially regarding late measures (see e.g., Godfroid 2020).

The presentation exemplifies how combining methods provides deeper insight into language variation and offers a more comprehensive understanding of both the production and processing of morphologically complex forms.

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

References

- Godfroid, Aline. 2020. *Eye Tracking in Second Language Acquisition and Bilingualism. A Research Synthesis and Methodological Guide*. New York: Routledge.
- Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka & Joost van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Koppel, Kristina & Jelena Kallas. 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 kirjallikus 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>.
- Lippus, Pärtel, Tanel Alumäe, Siim Orasmaa, Katrin Tsepelina & Liina Lindström. 2023a. Eesti Rahvusringhäälingu raadiosaadete korpus. [Estonian Public Broadcasting's Radio Corpus]. <https://doi.org/10.23673/re-441>.
- Lippus, Pärtel, Tanel Alumäe, Siim Orasmaa, Maarja-Liisa Pilvik & Liina Lindström. 2023b. Eesti taskuhäälingukorpus. [Estonian Podcast Corpus]. <https://doi.org/10.23673/re-445>.
- Pant, Annika. 2018. Asesõnade keegi, miski, kumbki käänevormide varieerumine eesti kirjakeeles. [The variation of case forms in pronouns keegi, miski, kumbki in written Estonian]. *Bakalaureusetöö*. <http://hdl.handle.net/10062/60560>.
- Pant, Annika. 2020. Pronoomenite keegi, miski, kumbki, ükski käänevormide kasutus tänapäeva eesti keeles. [The variation of case forms usage of pronouns keegi, miski, kumbki, ükski in modern Estonian]. *Magistritöö*. <http://hdl.handle.net/10062/68143>.
- Rull, Aado. 1917. Kaaslõpp -gi ja ta kidurad kaimud. [The clitic -gi and its lesser relatives]. *Eesti Kirjandus* 2. 82–88.
- Saareste, Albert. 1923. Kellegile, mingit, kuskil. [To someone, some, somewhere]. *Eesti Keel* 4. 116–117.
- Saareste, Andrus. 1936. Eesti õigekeelsuse päevaküsimustest. [On Current Issues in Estonian Orthography]. *Eesti Kirjandus* 12. 548–556.
- Saareste, Andrus. 1955. *Petit atlas des parlers estoniens. Väike eesti murdeatlas*. [A Small Estonian Dialect Atlas]. Uppsala: Almqvist & Wiksells.