

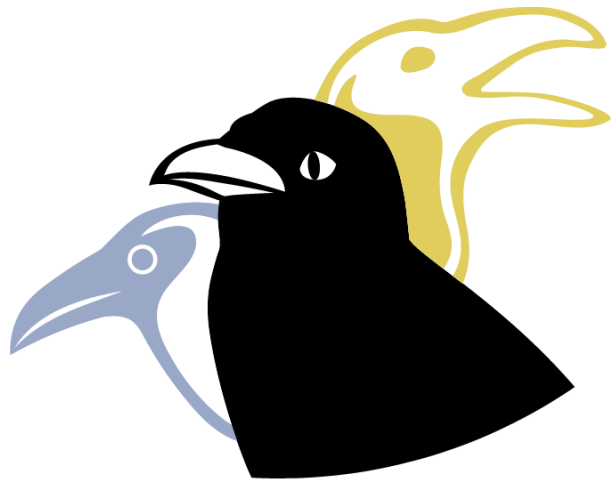
To reread or not to reread? Effects of frequency and standardization in the reading of morphological parallel forms in Estonian

Annika Kängsepp, Maarja-Liisa Pilvik, Kaidi Lõo, Külli Habicht & Liina Lindström
University of Tartu



SLE, Osnabrück
26 August 2026





VarEs

Morphosyntactic
variation in Estonian

Annika Kängsepp



Külli Habicht



Liina Lindström



Kaidi Lõo

Introduction

- Variation is an inherent property of language and occurs at all linguistic levels, including morphology.
- In languages with a standardized written variety, some variation is retained within the codified standard (e.g., parallel forms).

burnt ~ burned

Introduction

- Variation is an inherent property of language and occurs at all linguistic levels, including morphology.
- In languages with a standardized written variety, some variation is retained within the codified standard (e.g., parallel forms).

burnt ~ burned

- Other variants occur in language use but fall outside the codified standard. These have been described as **nonstandard forms/variants** (Trudgill 1999; Anderwald 2009), **vernacular forms/features** (Chambers 2004), **norm violations** (Hubers et al. 2020), etc.

*knew ~ *knowed*

Estonian situation

- Estonian nominal inflectional paradigm typically consists of 28 cells, representing **14 cases** and **2 numbers** (SG and PL).
- Morphological **parallel forms** in the same paradigm cell are relatively common, most frequently in plural forms (Viht & Habicht 2019; Aigro & Vihman 2023).

<i>põhjuse-i-d</i>	~	<i>põhjusi</i>	'reasons'
reason-PL-PAR		reason.PL.PAR	

- In addition to standard parallel forms, actual language use includes **nonstandard variants** occurring alongside standard ones.

<i>auto-sid</i>	~	* <i>auto-si</i>	'cars'
car-PL.PAR		car-PL.PAR	

<i>pere-t</i>	~	* <i>pere</i>	'family'
car.SG-PAR		car.SG.PAR	

Estonian situation

(Lindström, Risberg & Plado 2023)

- In Estonian language planning, **parallel forms and variation were discouraged** during most of the 20th century. Standard Estonian had a privileged status over dialects and only one standard form was preferred.
- At the end of the **1970s**, language planning became more democratic and greater attention was paid to actual language use, **more parallel forms were accepted into Standard Estonian**.
- Present day language planning is research based. However, **public attitudes** still remain strongly influenced by the **Standard Language Ideology** (see e.g., Algvere 2025; Pilvik et al. 2026).

Previous research

Previous research on morphological variation **in Estonian**:

- Corpus studies of standard-nonstandard morphological patterns
(e.g., Kängsepp 2024, 2025; Siiman 2019, 2023; Aigro, Todesk & Kängsepp 2026)
- Corpus studies of standard parallel forms
(e.g., Siiman 2019; Aigro & Vihman 2023, 2024)
- Acceptability of nonstandard forms
(e.g., Pant 2020; Pilvik et al. 2026)
- On-line production of parallel forms
(e.g., Pant 2020; Vihman et al. 2025)
- On-line comprehension of standard inflected forms
(e.g., Lõo et al. 2025a, 2025b)

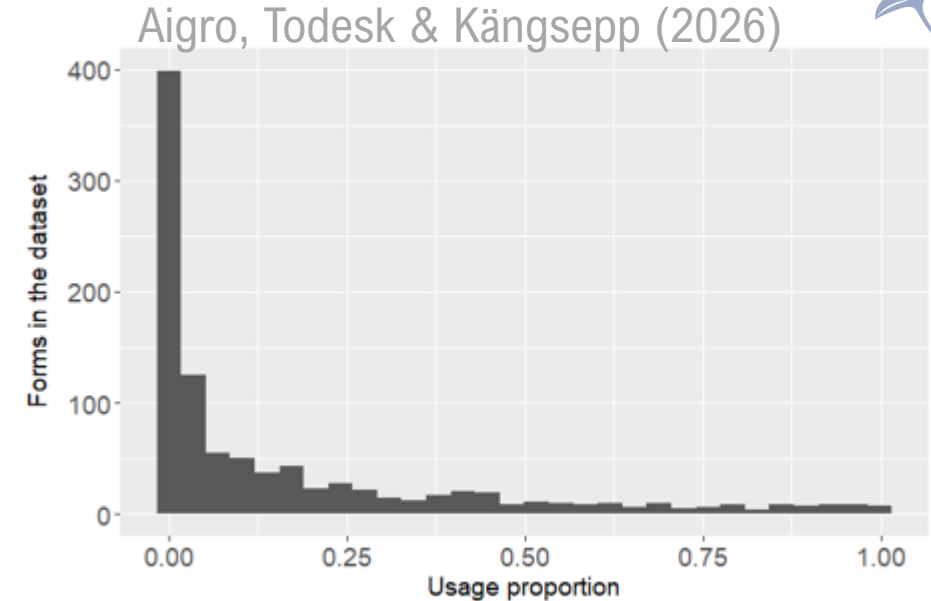
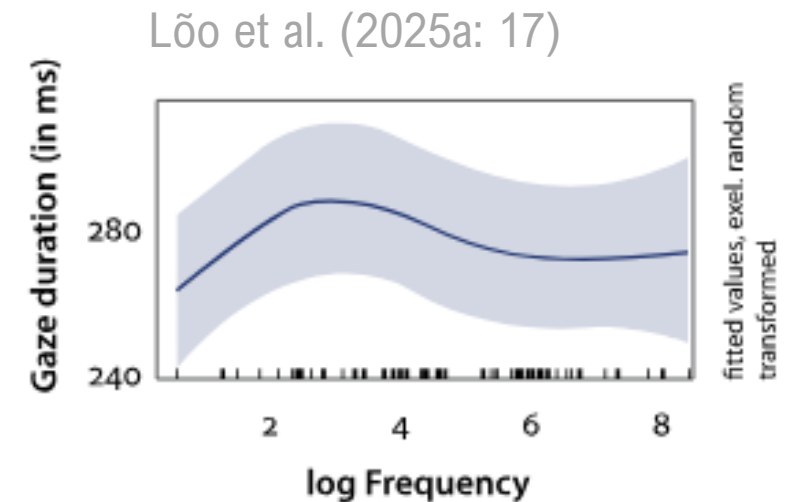


Figure 3. Histogram of the usage frequencies of nonstandard forms as a proportion of cell usage.



In this study

- We examine how **normative status** and **frequency** affect the reading of morphologically varying nominal forms in Estonian using **eye-tracking**. In addition, we explore the role of **individual and sociodemographic differences** in reading behaviour.

In this study

- We examine how **normative status** and **frequency** affect the reading of morphologically varying nominal forms in Estonian using **eye-tracking**. In addition, we explore the role of **individual and sociodemographic differences** in reading behaviour.
- We focus on 2 measures derived from eye movement data

(see e.g., Rayner 1998; Holmqvist et al. 2011):

Gaze Duration: the sum of all fixation durations on a target word form before the eyes leave it for the first time. Early processing measure, commonly interpreted as an index of lexical access / word recognition difficulty.

In this study

- We examine how **normative status** and **frequency** affect the reading of morphologically varying nominal forms in Estonian using **eye-tracking**. In addition, we explore the role of **individual and sociodemographic differences** in reading behaviour.
- We focus on 2 measures derived from eye movement data

(see e.g., Rayner 1998; Holmqvist et al. 2011):

Gaze Duration: the sum of all fixation durations on a target word form before the eyes leave it for the first time. Early processing measure, commonly interpreted as an index of lexical access / word recognition difficulty.

Total Dwell Time: the sum of all fixation durations on a target word form (incl. rereading durations). Late processing measure which may indicate greater interest, informativeness, or processing difficulty.

Previous research

Previous reading studies about the processing of standard-nonstandard constructional variation have found that

- **norm violations are more difficult to process** than grammatical constructions (Squires 2016; Hubers et al. 2020; Cos & Hubers 2022),
- norm violations pattern with ungrammatical constructions in **late measures of reading** (Hubers et al. 2020; Cos & Hubers 2022),
- **frequency has a facilitating effect** on reading, independent of the standard-nonstandard status of the form (Squires 2016),
- **individual differences** modulate the effect of norm violations in reading tasks (Hubers et al. 2020; Cos & Hubers 2022).

Research questions and hypotheses

- Q1:** Are **standard forms read faster** than non-standard forms?
- Q2:** Are reading time differences reflected in **early or late measures** of reading?
- Q3:** Does **the effect of frequency** on reading times depend on the status of the form read?
- Q4:** Do **individual differences or sociodemographic characteristics** affect reading time differences between standard and non-standard forms?
- Q5:** Do the reading time differences between standard-nonstandard forms **differ from those between standard parallel forms**?

Research questions and hypotheses

- H1:** Standard forms **are read faster** than non-standard forms.
- Q2:** Are reading time differences reflected in **early or late measures** of reading?
- Q3:** Does **the effect of frequency** on reading times depend on the status of the form read?
- Q4:** Do **individual differences or sociodemographic characteristics** affect reading time differences between standard and non-standard forms?
- Q5:** Do the reading time differences between standard-nonstandard forms **differ from those between standard parallel forms**?

Research questions and hypotheses

- H1:** Standard forms **are read faster** than non-standard forms.
- H2:** Reading time differences are reflected in **late measures** of reading.
- Q3:** Does **the effect of frequency** on reading times depend on the status of the form read?
- Q4:** Do **individual differences or sociodemographic characteristics** affect reading time differences between standard and non-standard forms?
- Q5:** Do the reading time differences between standard-nonstandard forms **differ from those between standard parallel forms**?

Research questions and hypotheses

- H1:** Standard forms **are read faster** than non-standard forms.
- H2:** Reading time differences are reflected in **late measures** of reading.
- H3:** The effect of frequency on reading times **does not depend** on the status of the form read.
- Q4:** Do **individual differences or sociodemographic characteristics** affect reading time differences between standard and non-standard forms?
- Q5:** Do the reading time differences between standard-nonstandard forms **differ from those between standard parallel forms**?

Research questions and hypotheses

- H1:** Standard forms **are read faster** than non-standard forms.
- H2:** Reading time differences are reflected in **late measures** of reading.
- H3:** The effect of frequency on reading times **does not depend** on the status of the form read.
- H4:** Individual differences and sociodemographic characteristics **affect** reading time differences between standard and non-standard forms.
- Q5:** Do the reading time differences between standard-nonstandard forms **differ from those between standard parallel forms**?

Research questions and hypotheses

- H1:** Standard forms **are read faster** than non-standard forms.
- H2:** Reading time differences are reflected in **late measures** of reading.
- H3:** The effect of frequency on reading times **does not depend** on the status of the form read.
- H4:** Individual differences and sociodemographic characteristics **affect** reading time differences between standard and non-standard forms.
- H5:** Reading time differences between standard-nonstandard forms **do not differ from those between standard parallel forms**.

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'standard dominant'

imet ~ *ime* ('miracle' SG.PAR)

'equal'

peret ~ *pere* ('family' SG.PAR)

'nonstandard dominant'

hüvet ~ *hüve* ('benefit' SG.PAR)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'older dominant'

vaidlusi ~ *vaidluseid* ('argument' PL.PAR)

'equal'

põhjusi ~ *põhjuseid* ('reason' PL.PAR)

'newer dominant'

jäätisi ~ *jäätiseid* ('ice cream' PL.PAR)

52 fillers

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'standard dominant'
imet ~ *ime* ('miracle' SG.PAR)

'equal'
peret ~ *pere* ('family' SG.PAR)

'nonstandard dominant'
hüvet ~ *hüve* ('benefit' SG.PAR)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'older dominant'
vaidlusi ~ *vaidluseid* ('argument' PL.PAR)

'equal'
põhjusi ~ *põhjuseid* ('reason' PL.PAR)

'newer dominant'
jäätisi ~ *jäätiseid* ('ice cream' PL.PAR)

52 fillers

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'standard dominant'

imet ~ *ime* ('miracle' SG.PAR)

'equal'

peret ~ *pere* ('family' SG.PAR)

'nonstandard dominant'

hüvet ~ *hüve* ('benefit' SG.PAR)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'older dominant'

vaidlusi ~ *vaidluseid* ('argument' PL.PAR)

'equal'

põhjusi ~ *põhjuseid* ('reason' PL.PAR)

'newer dominant'

jäätisi ~ *jäätiseid* ('ice cream' PL.PAR)

52 fillers

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'standard dominant'
imet ~ *ime* ('miracle' SG.PAR)

'equal'
peret ~ *pere* ('family' SG.PAR)

'nonstandard dominant'
hüvet ~ *hüve* ('benefit' SG.PAR)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

'older dominant'
vaidlusi ~ *vaidluseid* ('argument' PL.PAR)

'equal'
põhjusi ~ *põhjuseid* ('reason' PL.PAR)

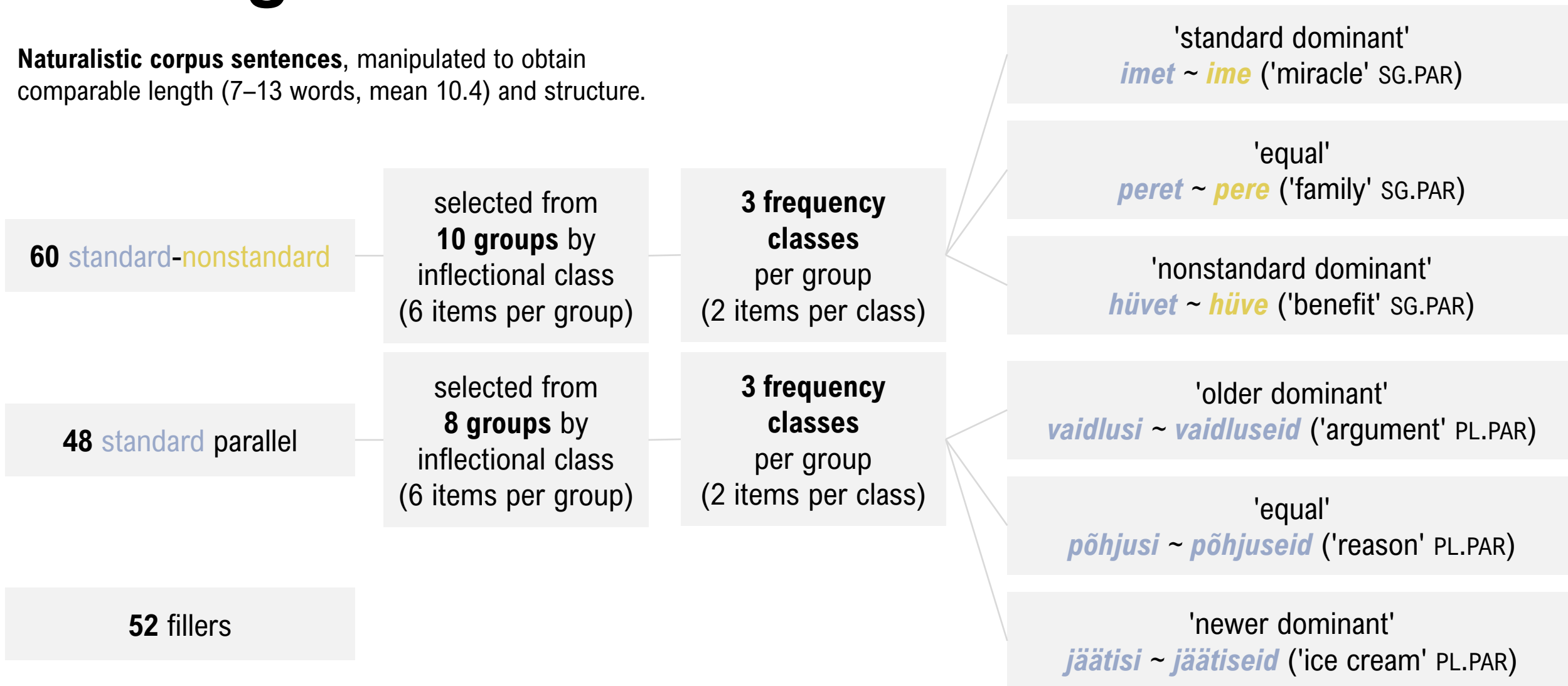
'newer dominant'
jäätisi ~ *jäätiseid* ('ice cream' PL.PAR)

52 fillers

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

52 fillers

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



'standard dominant'
imet ('miracle' SG.PAR)

'equal'
pere ('family' SG.PAR)

'nonstandard dominant'
hüvet ('benefit' SG.PAR)

'older dominant'
vaidluseid ('argument' PL.PAR)

'equal'
põhjusi ('reason' PL.PAR)

'newer dominant'
jäätiseid ('ice cream' PL.PAR)

Design

Naturalistic corpus sentences, manipulated to obtain comparable length (7–13 words, mean 10.4) and structure.

60 standard-nonstandard

selected from
10 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

48 standard parallel

selected from
8 groups by
inflectional class
(6 items per group)

**3 frequency
classes**
per group
(2 items per class)

52 fillers

Based on frequencies from
the **Estonian National Corpus** (2023),
3.8 billion tokens (Koppel et al. 2023)



'standard dominant'
ime ('miracle' SG.PAR)

peret 'equal'
(('family' SG.PAR)

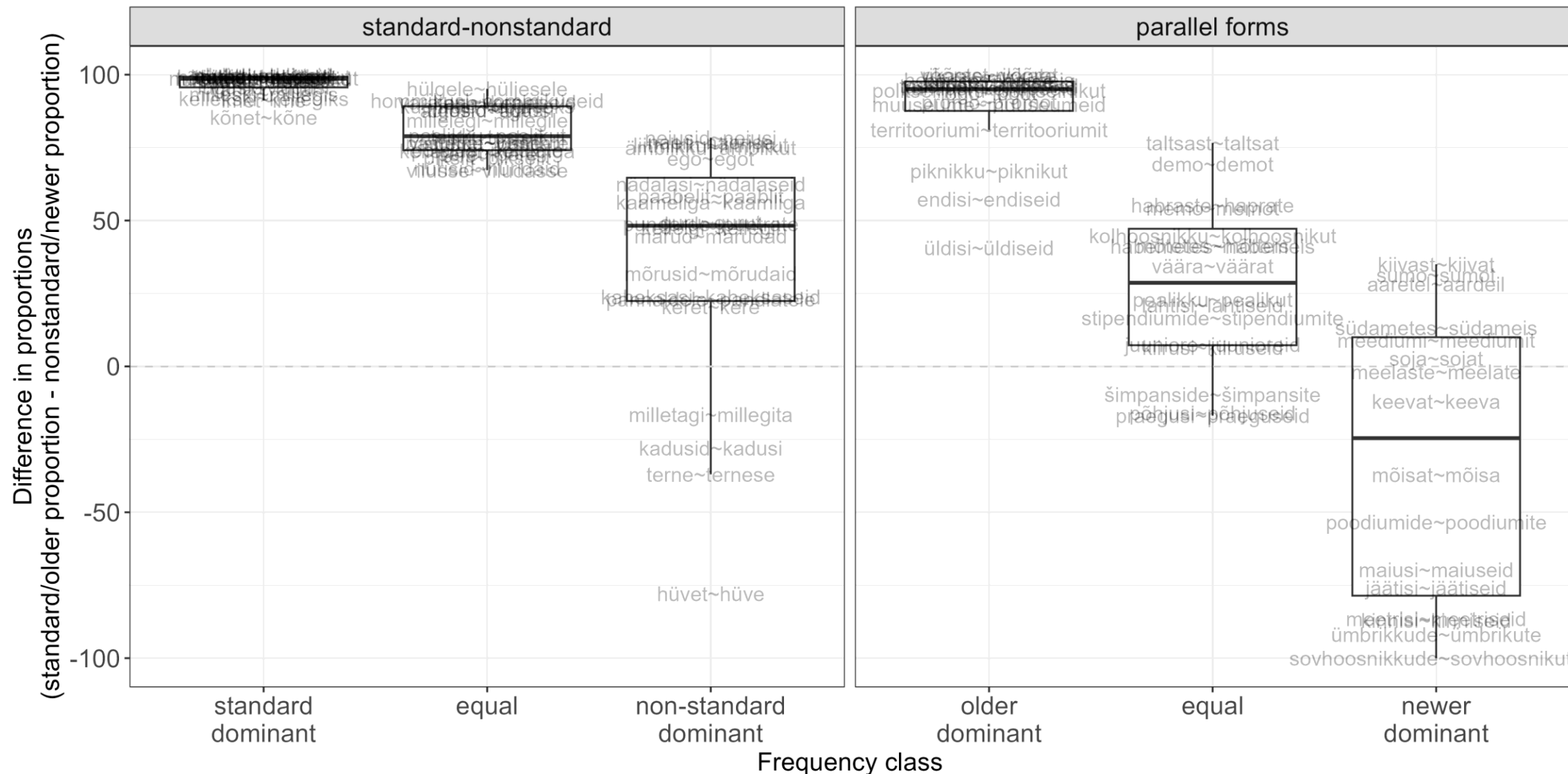
'nonstandard dominant'
hüve ('benefit' SG.PAR)

vaidlusi 'older dominant'
(('argument' PL.PAR)

'equal'
põhjuseid ('reason' PL.PAR)

jäätisi 'newer dominant'
(('ice cream' PL.PAR)

Distribution of the stimuli



Participants

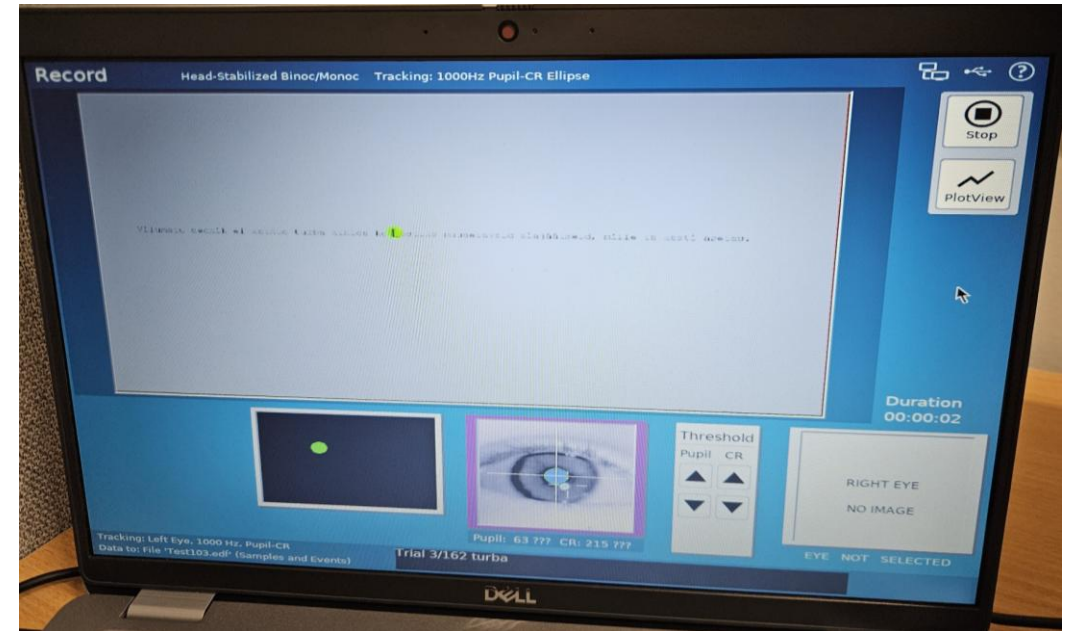
84 L1 Estonian speakers were recruited

- aged **18–59** years
($M = 34.96$, $SD = 11.98$)
- **47 F**, **34 M**, **3** other
- **51** with higher education (BA or higher),
33 with lower

Before the reading task, participants completed
a short **background questionnaire**

and **the Big Five Inventory (BFI-44)**

designed to measure personality traits (Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness).



Analysis

- Generalized Additive Mixed Models using *mgcv* package in R (Wood 2017)

DV ~ **Variant** +
 Log frequency of word form +
 Log frequency of alternative word form +
 Relative Frequency Difference +
 Word Length +
 Relative Position in Sentence +
 Participant Gender +
 Participant Age +
 Participant Education +
 Participant Reading Habits for Work or School +
 Participant Reading Habits in Free Time +
 Trial Index +
 Random Intercept for Participant +
 Random Intercept for Item

Analysis

- Generalized Additive Mixed Models using *mgcv* package in R (Wood 2017)

DV ~	Variant + Log frequency of word form + Log frequency of alternative word form + Relative Frequency Difference + Word Length + Relative Position in Sentence + Participant Gender + Participant Age + Participant Education + Participant Reading Habits for Work or School + Participant Reading Habits in Free Time + Trial Index + <i>Random Intercept for Participant +</i> <i>Random Intercept for Item</i>	Tested interactions between Variant * Frequency Gender Age Education Reading Frequency * Alternative form frequency
------	--	--

Analysis

- Generalized Additive Mixed Models using *mgcv* package in R (Wood 2017)

DV ~	Variant +	Tested random slopes	
	Log frequency of word form +		
	Log frequency of alternative word form +	Variant (contr)	*
	Relative Frequency Difference +		Participant
	Word Length +		Item
	Relative Position in Sentence +		
	Participant Gender +		
	Participant Age +		
	Participant Education +		
	Participant Reading Habits for Work or School +		
	Participant Reading Habits in Free Time +		
	Trial Index +		
	<i>Random Intercept for Participant +</i>		
	<i>Random Intercept for Item</i>		

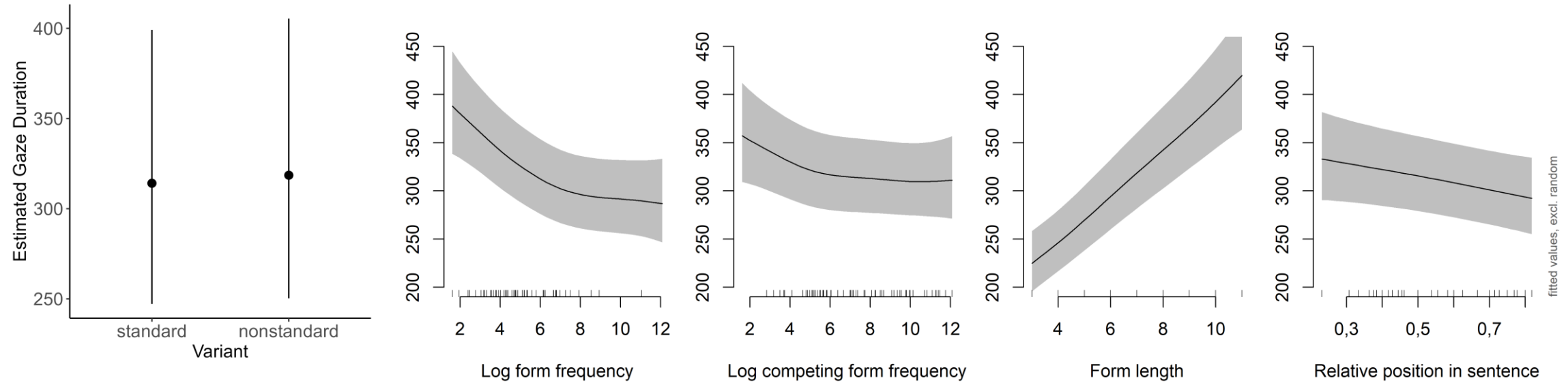
Analysis

- Generalized Additive Mixed Models using *mgcv* package in R (Wood 2017)

DV ~	Variant + Log frequency of word form + Log frequency of alternative word form + Relative Frequency Difference + Word Length + Relative Position in Sentence + Participant Gender + Participant Age + Participant Education + Participant Reading Habits for Work or School + Participant Reading Habits in Free Time + Trial Index + <i>Random Intercept for Participant</i> + <i>Random Intercept for Item</i>	Tested BFI scores and interactions between Variant * Agreeableness Conscientiousness Extraversion Neuroticism Openness
------	--	--

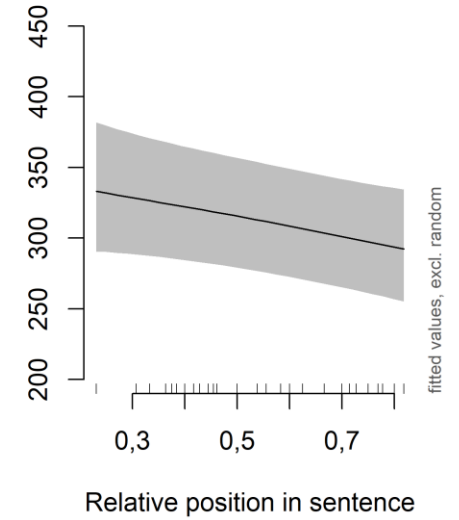
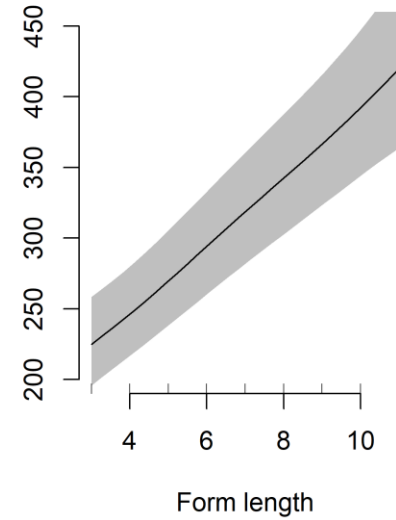
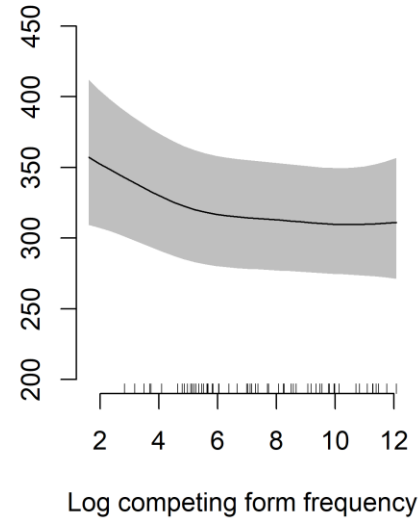
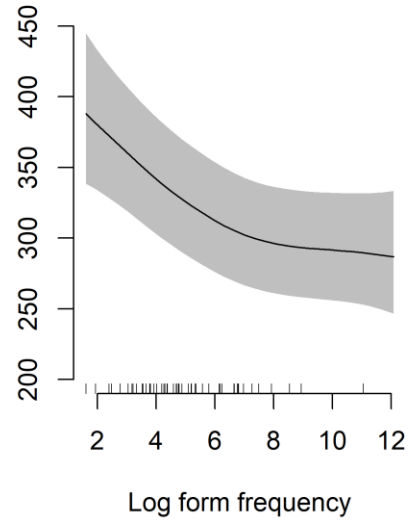
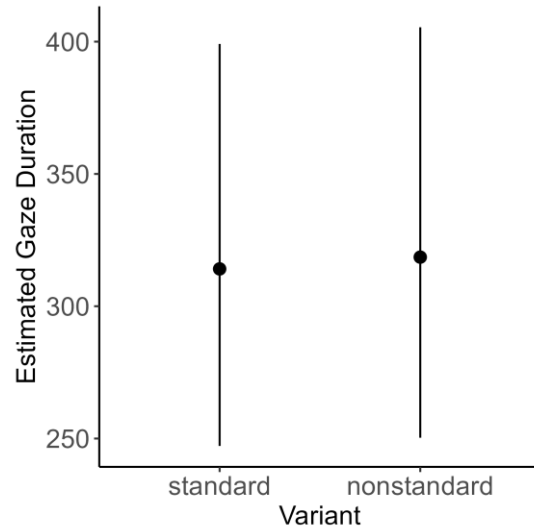
Results: Gaze Duration

S-NS

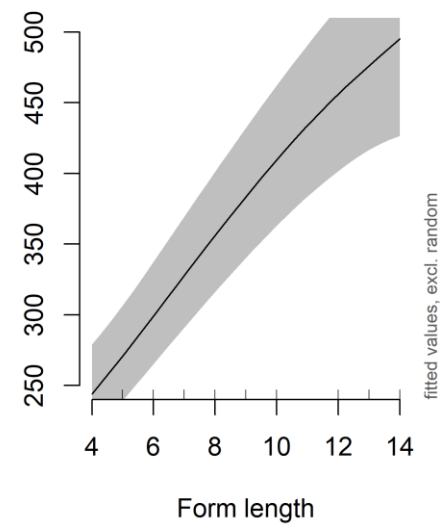
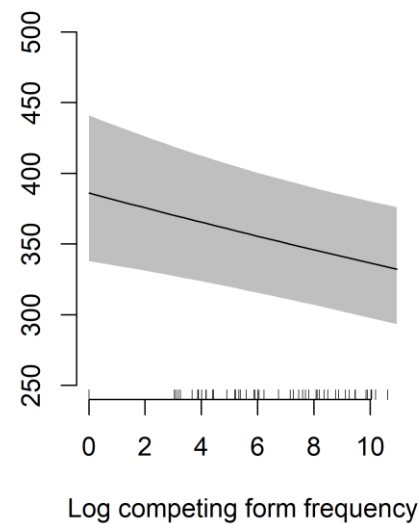
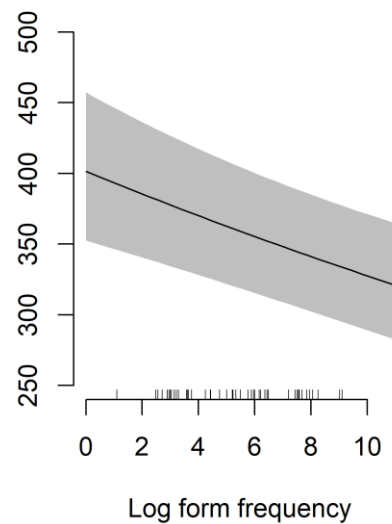
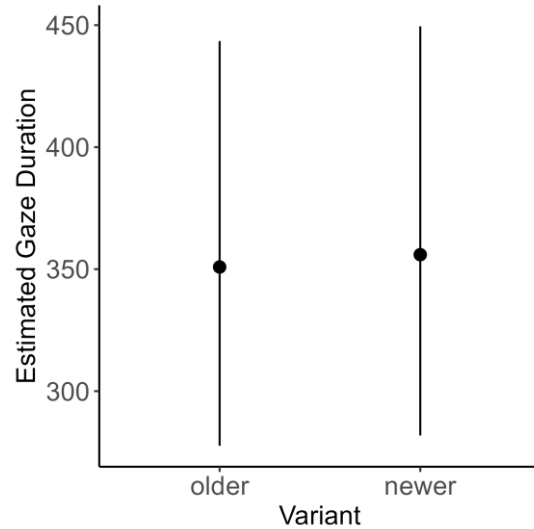


Results: Gaze Duration

S-NS

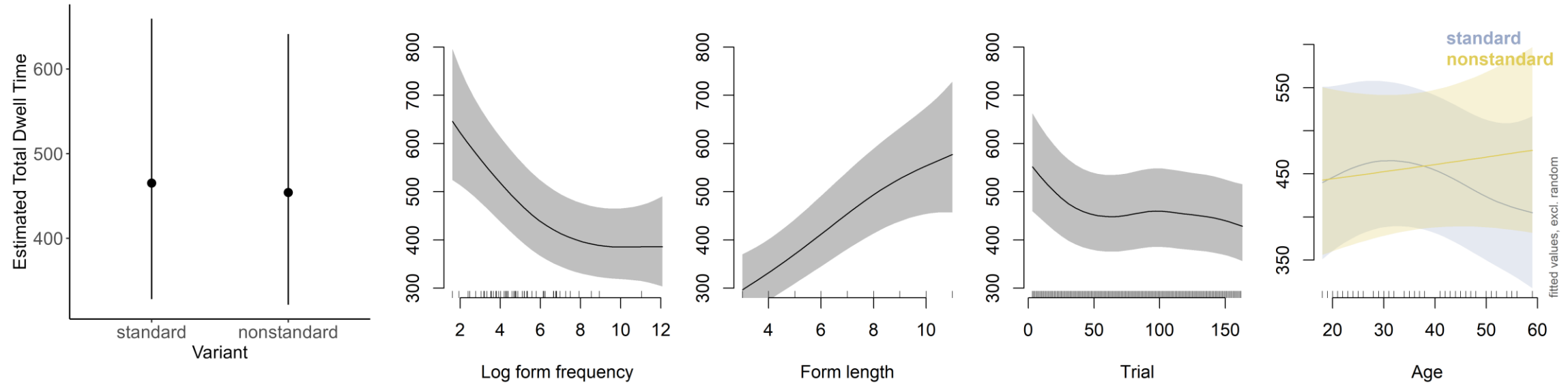


S-S



Results: Total Dwell Time

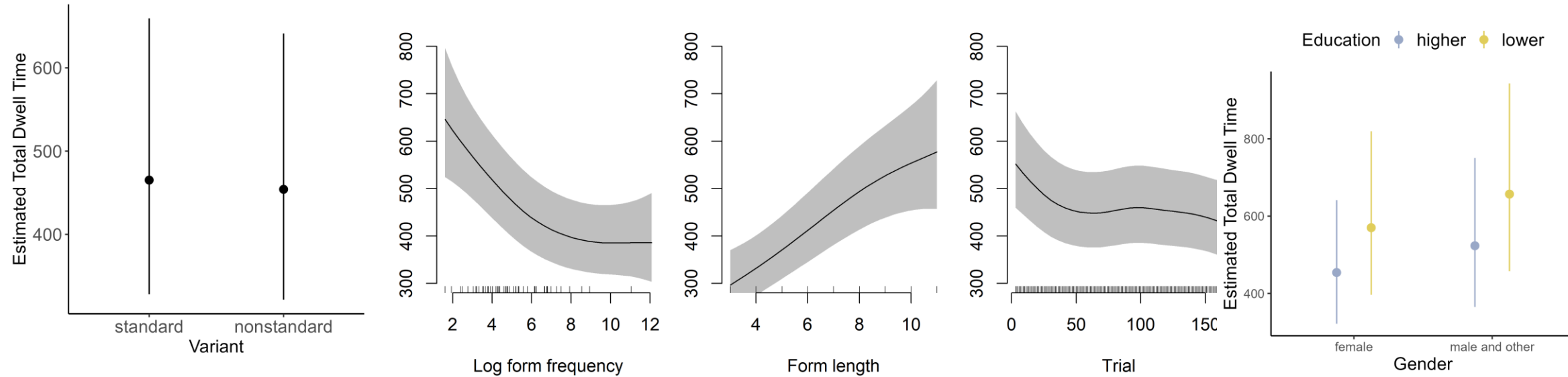
S-NS



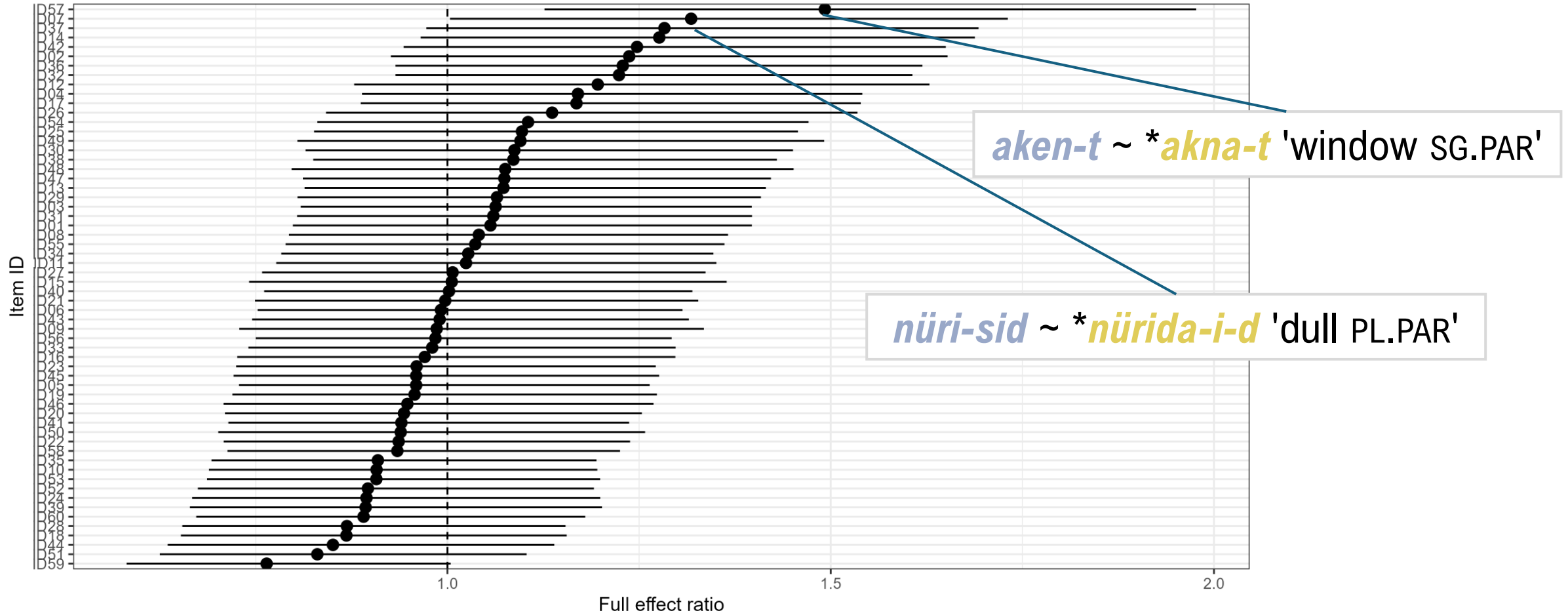
fitted values, excl. random

Results: Total Dwell Time

S-NS

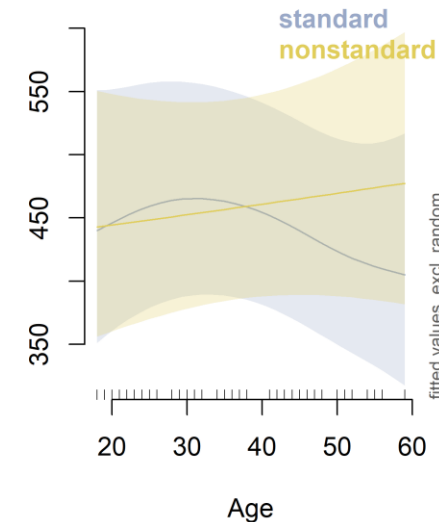
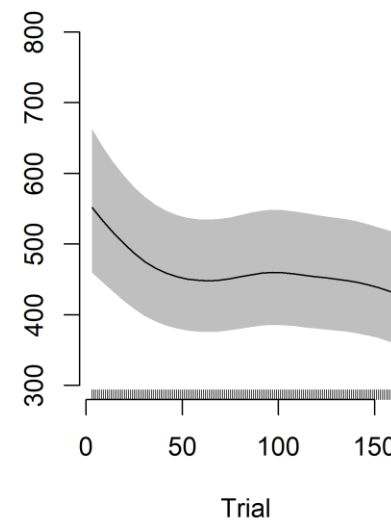
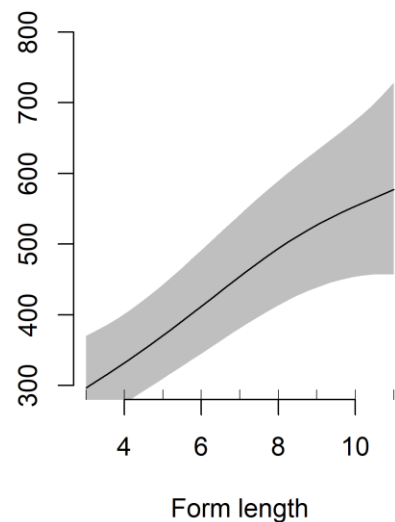
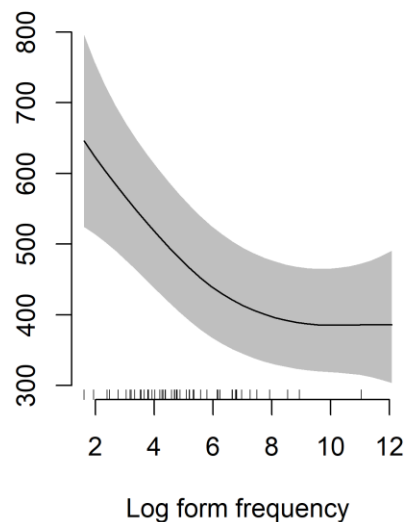
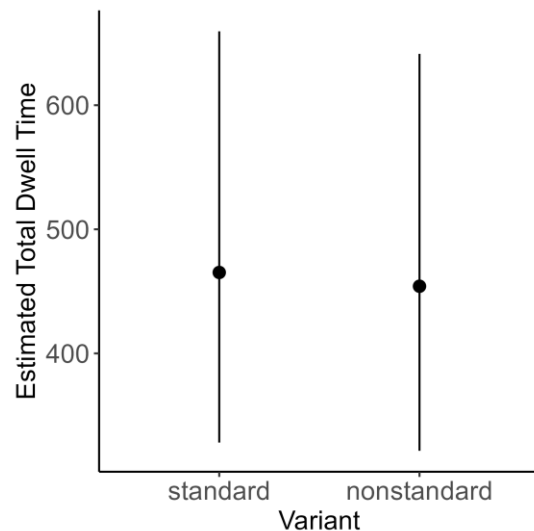


Results: Total Dwell Time

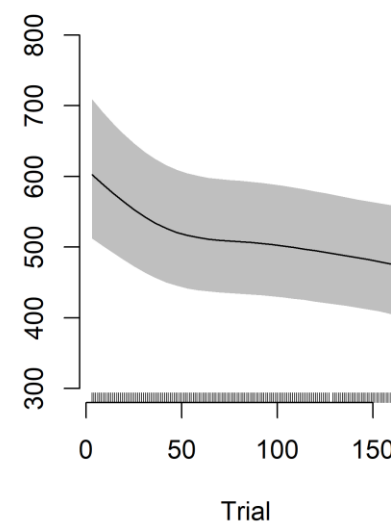
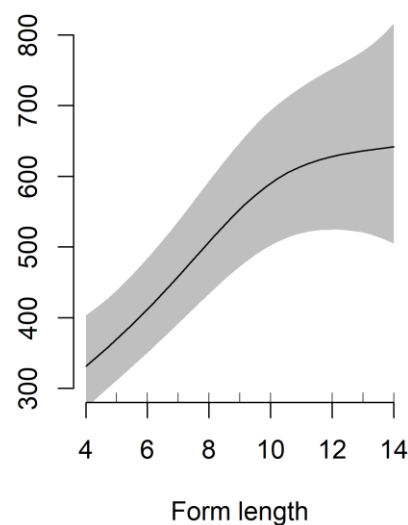
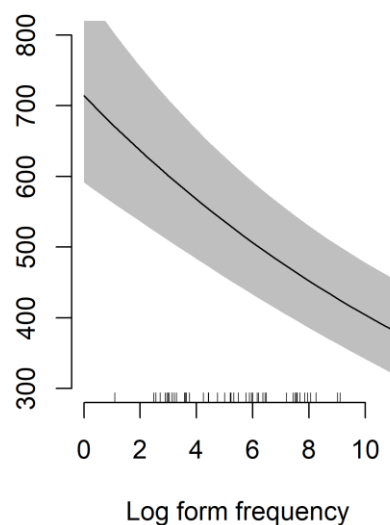
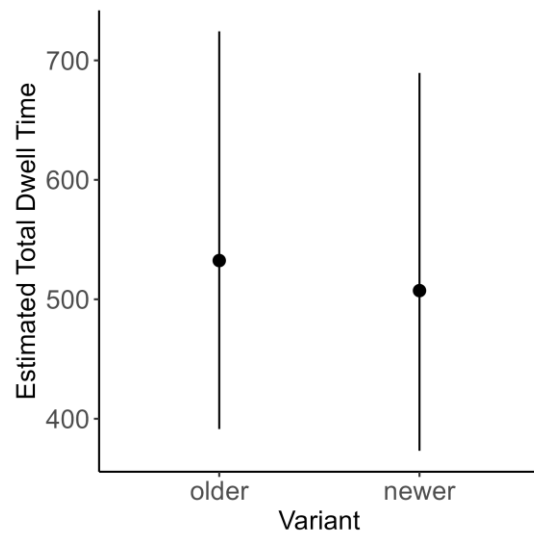


Results: Total Dwell Time

S-NS

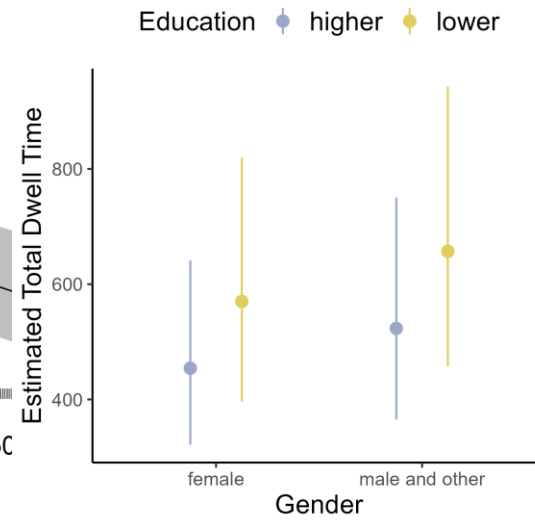
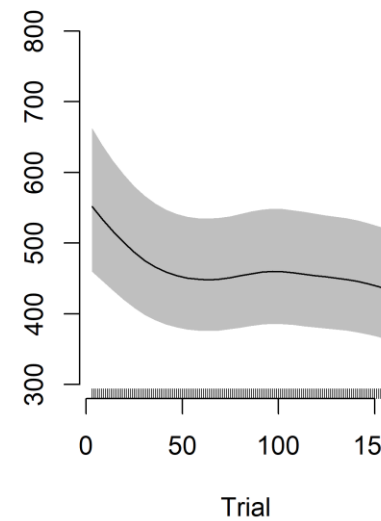
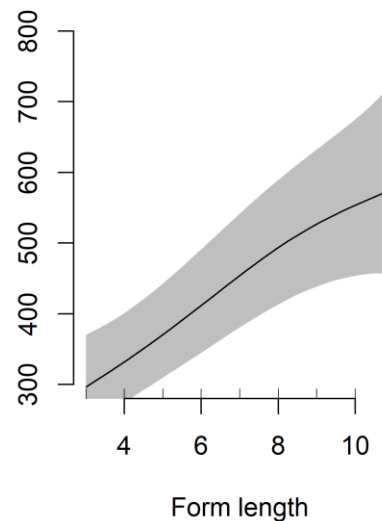
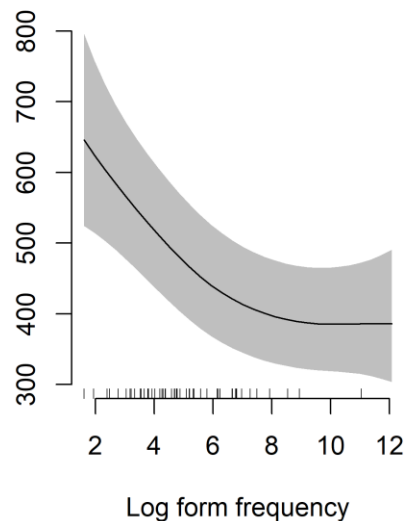
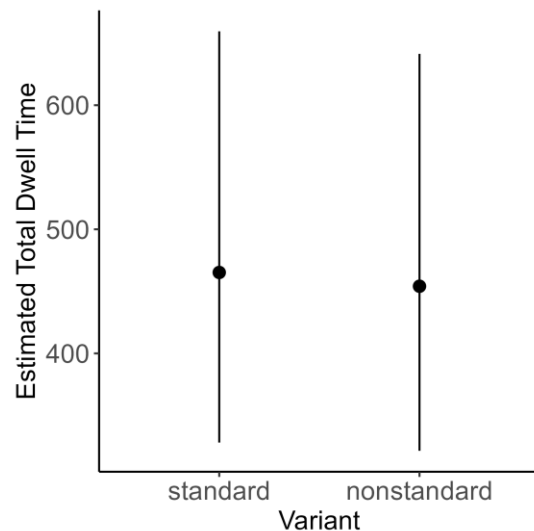


S-S

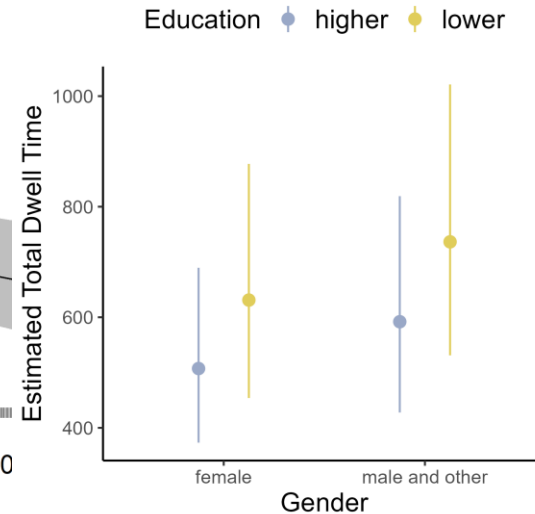
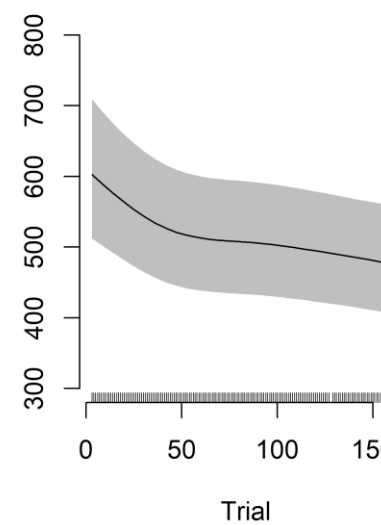
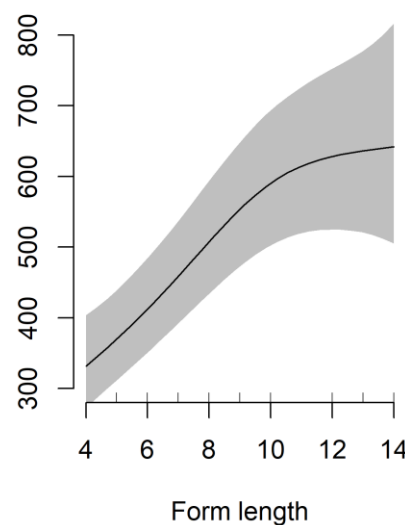
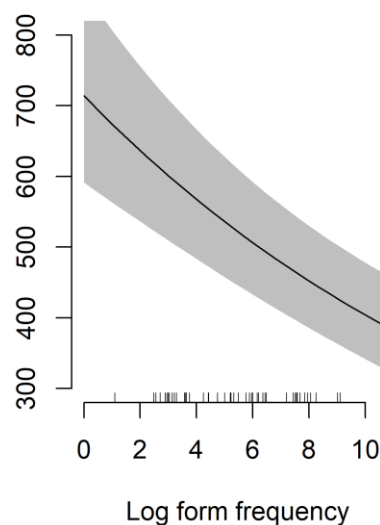
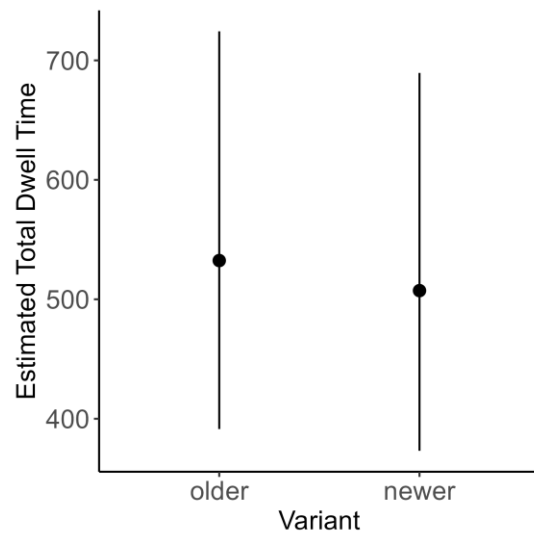


Results: Total Dwell Time

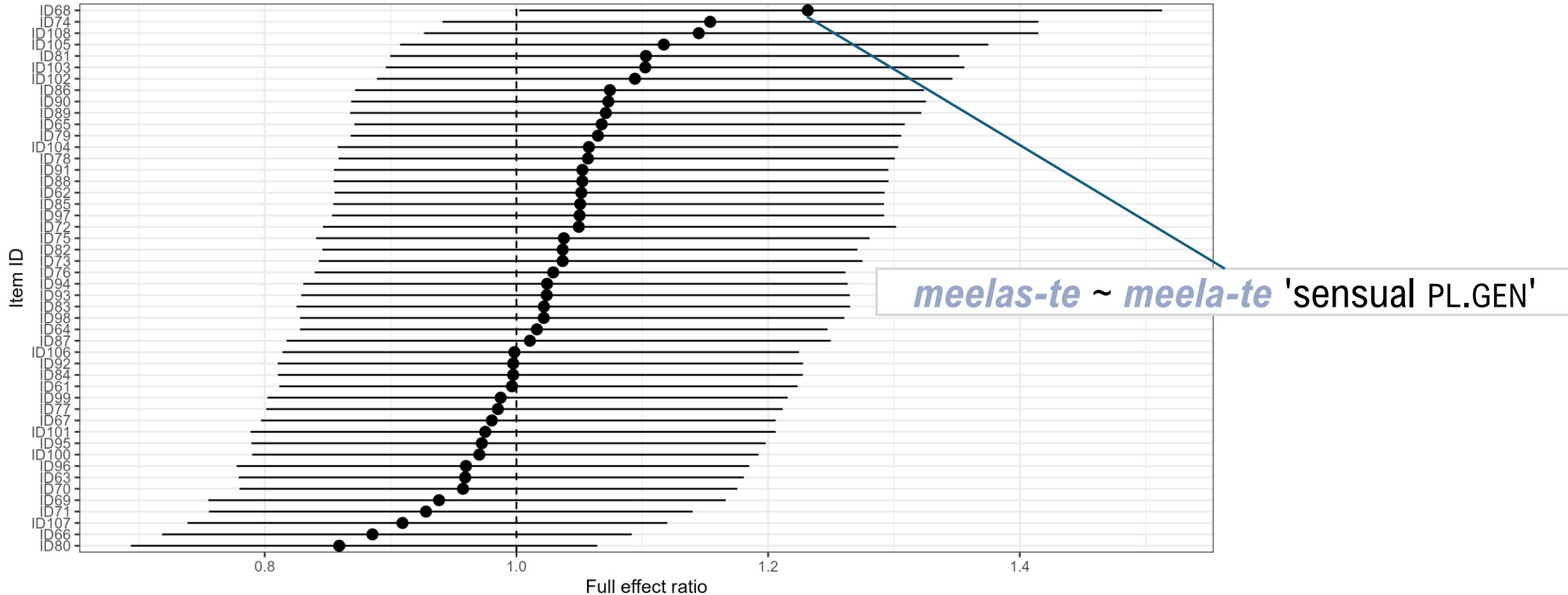
S-NS



S-S



Results: Total Dwell Time



Summary

No overall reading cost for nonstandard morphological forms.

- Standard forms were not read faster than nonstandard forms.
- Higher whole-word frequency was associated with faster reading, regardless of the normative status of the form.
- Individual differences emerged in late stages of reading:
 - In total reading time, higher-educated and female readers were faster than lower-educated and non-female readers, regardless of the form status.
 - Older participants showed slightly longer total reading times for nonstandard forms.
- Comparisons between standard parallel forms showed patterns similar to those observed for standard–nonstandard form pairs.

References

- Aigro, Mari; Todesk, Triin; Kängsepp, Annika 2026.** Increasing transparency: Factors in the usage of nonstandard inflections by Estonian adults. – *Journal of Linguistics*, to appear.
- Aigro, Mari; Vihman, Virve-Anneli 2023.** Realised overabundance in Estonian noun paradigms: A corpus study. – *Word Structure* 16(2-3), 154–175. <https://doi.org/10.3366/word.2023.0227>
- Aigro, Mari, Vihman, Virve-Anneli 2024.** Preferences in the use of overabundance: Predictors of lexical bias in Estonian. *Cognitive Linguistics* 35(2), 289–312. <http://doi.org/10.1515/cog-2023-0035>
- Algvere, Kristel 2025.** Adult L1 users' attitudes towards their proficiency in written Estonian as evidence of deep-rooted language ideologies. – *ESUKA / JEFUL* 16(2), 5–50. <https://doi.org/10.12697/jeful.2025.16.2.01>
- Anderwald, Lieselotte 2009.** *The Morphology of English Dialects: Verb-Formation in Non-standard English*. Cambridge University Press.
- Chambers, J. K. 2004.** Dynamic typology and vernacular universals. In Berndt Kortmann (Ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Mouton de Gruyter, 127–145.
- Cos, Floris; Hubers, Ferdy 2022.** Variation in the processing of grammatical norm violations: Between-participant and between-construction differences. – *Linguistics in the Netherlands* 39, 21–38. <https://doi.org/10.1075/avt.00059.cos>
- Hubers, Ferdy; Redl, Theresa; de Vos, Hugo; Reinartz, Lukas; de Hoop, Helen 2020.** Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking. – *Frontiers in Psychology* 11:186. <https://doi.org/10.3389/fpsyg.2020.00186>
- Holmqvist, Kenneth; Nyström, Marcus; Andersson, Richard; Dewhurst, Richard; Jarodzka, Halszka; Van de Weijer, Joost 2011.** *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

References

- Kängsepp, Annika 2024.** Käänevormide varieerumine ja seda mõjutavad tegurid kirjalikus keeles indefiniitpronoomeni *keegi* näitel. – *Keel ja Kirjandus* 11, 1016–1037. <https://doi.org/10.54013/kk803a3>
- Kängsepp, Annika 2025.** Indefiniitpronoomenite *keegi* ja *miski* käänevormide varieerumine suulises keeles. – *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics* 21, 121–139. <https://doi.org/10.5128/ERYa21.07>
- Lindström, Liina; Risberg, Lydia; Plado, Helen 2023.** Language ideologies and beliefs about language in Estonia and Estonian language planning. – *ESUKA / JEFUL* 14(1), 7–48. <https://doi.org/10.12697/jeful.2023.14.1.01>
- Lõo, Kaidi; Bertram, Raymond; Kuperman, Victor 2025a.** Long-lag morphological priming and inflectional paradigm size effects in Estonian and Finnish text reading. In Lori Buchanan & Roberto G. de Almeida (Eds.), *The Representation and Processing of Morphologically Complex Words*. (*The Mental Lexicon* 19:2.) John Benjamins, 253–284. <https://doi.org/10.1075/ml.24035.loo>
- Lõo, Kaidi; Tavast, Arvi; Heitmeier, Maria; Baayen, R. Harald 2025b.** An experimental and computational study of an Estonian single-person word naming. arXiv: <https://doi.org/10.48550/arXiv.2509.03143>
- Pant, Annika 2020.** *Pronoomenite keegi, miski, kumbki, ükski käänevormide kasutus tänapäeva eesti keeles*. MA Thesis. University of Tartu.
- Pilvik, Maarja-Liisa; Lindström, Liina; Algvere, Kristel; Risberg, Lydia; Plado, Helen; Kängsepp, Annika 2026.** Milline keelekasutus eestlasi häirib? – *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics* 22, 127–144. <https://doi.org/10.5128/ERYa22.08>

References

- Rayner, Keith 1998.** Eye movements in reading and information processing: 20 years of research. – *Psychological Bulletin* 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Siiman, Ann 2019.** *Vormikasutuse varieerumine ning selle põhjused osastava ja sisseütleva käände näitel.* (Dissertationes philologiae estonicae Universitatis Tartuensis 45.) Tartu Ülikooli Kirjastus.
- Siiman, Ann 2023.** Normingu määramisest: lühikese ülivõrde varieerumise juhtumiuuring. *ESUKA / JEFUL* 14(1), 175–225. <https://doi.org/10.12697/jeful.2023.14.1.06>
- Squires, Lauren 2016.** Processing grammatical differences: Perceiving versus Noticing. In Anna M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research*. Cambridge University Press, 80–103. <https://doi.org/10.1017/CBO9781139680448.006>
- Trudgill, Peter 1999.** Standard English: What it isn't. In Tony Bex, Richard J. Watts (Eds.), *Standard English: The Widening Debate*. Routledge, 117–128.
- Vihman, Virve-Anneli; Hržica, Gordana; Aigro, Mari; Košutar, Sara; Bošnjak; Botica, Tomislava 2025.** Acquisition of morphological variation: An elicitation experiment on children's production of parallel forms in Croatian and Estonian. – *Journal of Child Language* 53(3), 539–473. <https://doi.org/10.1017/S0305000925000017>
- Viht, Annika; Habicht, Külli 2019.** *Eesti keele sõnamuutmine.* (Eesti keele varamu 4.) Tartu Ülikooli Kirjastus.
- Wood, Simon 2017.** *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>

Thank you!

