



UNIVERSITY OF TARTU

# Estonian language variation: historical background, language attitudes, and processing variation

Liina Lindström  
Maarja-Liisa Pilvik  
Kaidi Lõo

University of Tartu  
15.10.2025



# Outline

- Estonia and Estonian
- Some current sociolinguistic topics:
  - Russian minority and transition to Estonian school
  - South Estonian minority language
- Standard Language Ideology in Estonia
- Project: Morphosyntactic Variation in Estonian
- 2 case studies
  - What do Estonian speakers find annoying in the use of Estonian?
  - Does morphological variation affect reading?

# Why are we here?

In 2022, the Department of Estonian of the University of Tartu received inheritance from **Kadri Rõuk**, in the name of her parents Gerda and Nikolai Rõuk.

Study on Canadian Estonian heritage speaker's speech

- Fieldwork in the Alberta region (Edmonton, Stettler, Calgary, Canmore)
- Collaboration with the Alberta Estonian Heritage Society and University of Alberta in Edmonton (Anja Arnhold, Juhani Järvikivi)
- Special thanks to **Helgi Leesment** and **Kelly Schuler-McDonald**!



# Estonian

A **Uralic language**, spoken in the Republic of Estonia

Population of Estonia:  
1.3 million speakers

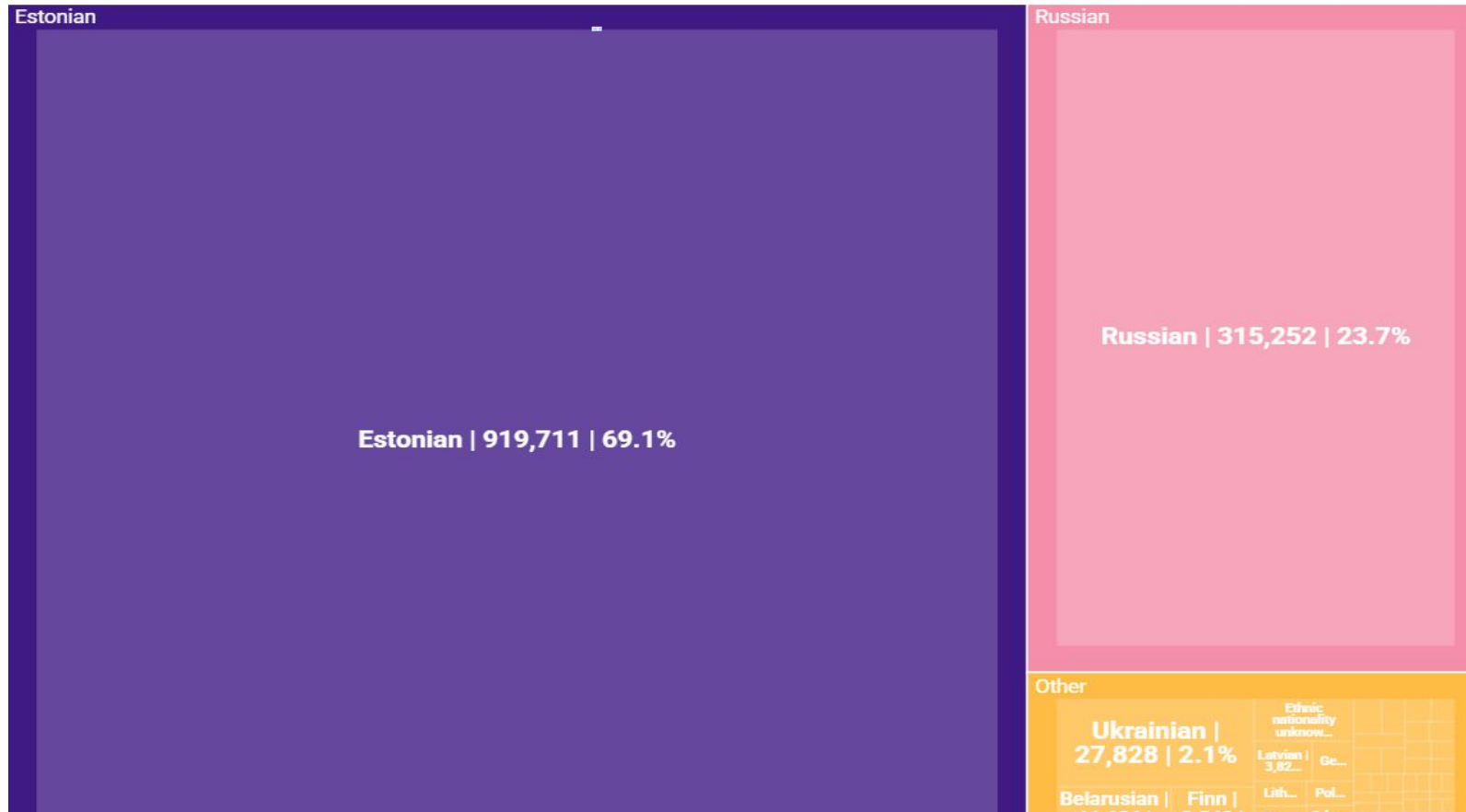
Official language:  
Estonian

Speakers of Estonian:  
about 0.9 millions in Estonia  
+ 0.1-0.2 millions abroad

Big Russian minority from the Soviet period



# Census 2021: population distribution by ethnic nationality



Estonian is spoken by 84% of the population of Estonia

67% of people speak it as **L1**

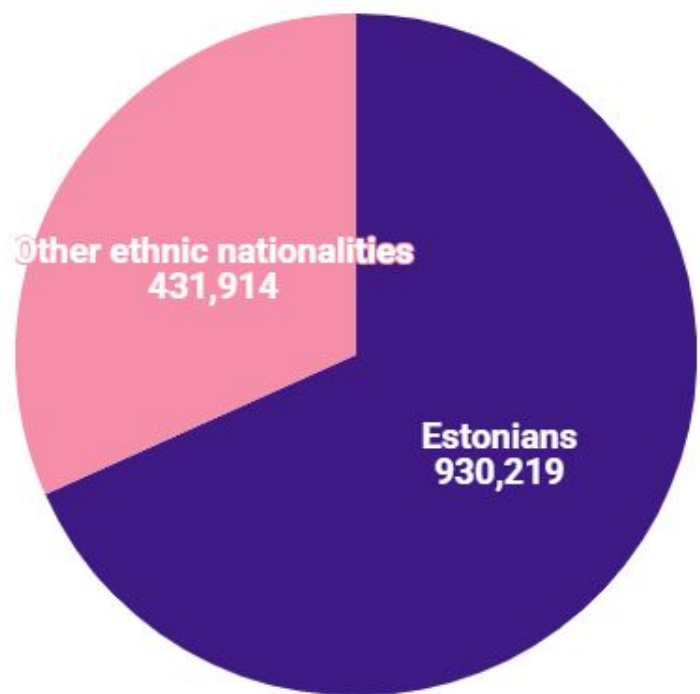
The proportion of Estonian **L2** speakers has steadily increased:

|       |     |
|-------|-----|
| 2000: | 12% |
| 2011: | 14% |
| 2021: | 17% |

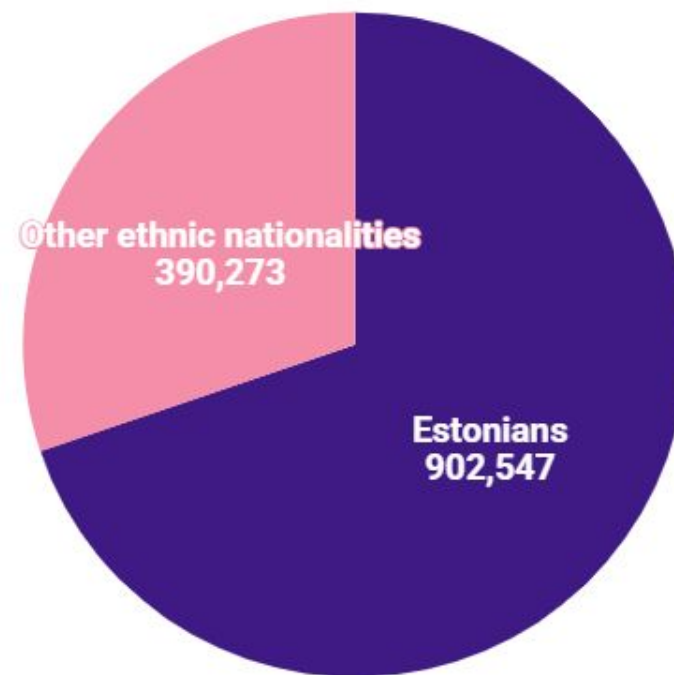
# Brief historical timeline

|                   |                                                                                                                                  |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------|
| <b>1918</b>       | Independence                                                                                                                     |
| <b>1918-1920</b>  | Independence war (against Soviet Russia)                                                                                         |
| <b>1940</b>       | Soviet troops march in, Estonia incorporated into Soviet Union                                                                   |
| <b>1941-1944</b>  | German troops occupy Estonian territory                                                                                          |
| <b>1944 Sept</b>  | Estonia reannexed by the Soviet Union; more than 80.000 Estonians escaped to the West (Sweden, Germany → Canada, Australia, USA) |
| <b>1941, 1949</b> | More than 30,000 Estonians deported to Siberia and Central Asia                                                                  |
| <b>1991</b>       | Reindependence                                                                                                                   |
| <b>2004</b>       | Estonia joined NATO and European Union                                                                                           |

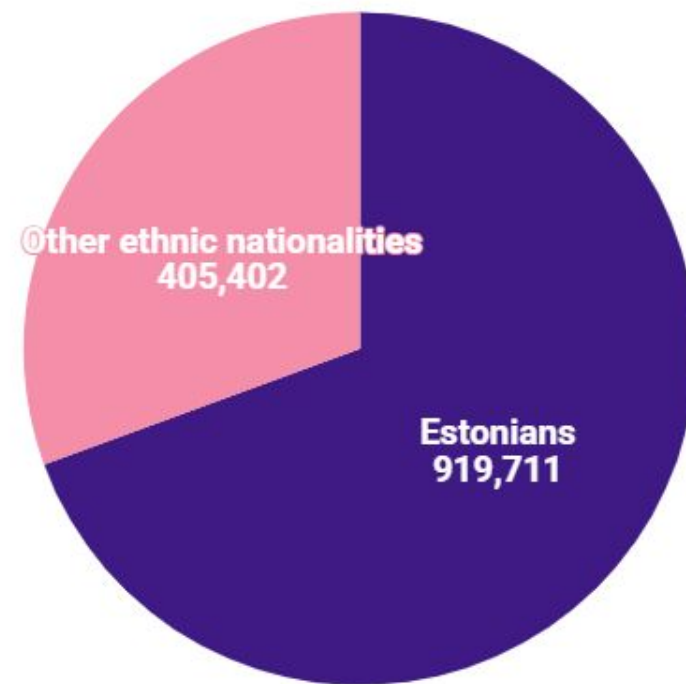
2000



2011



2021

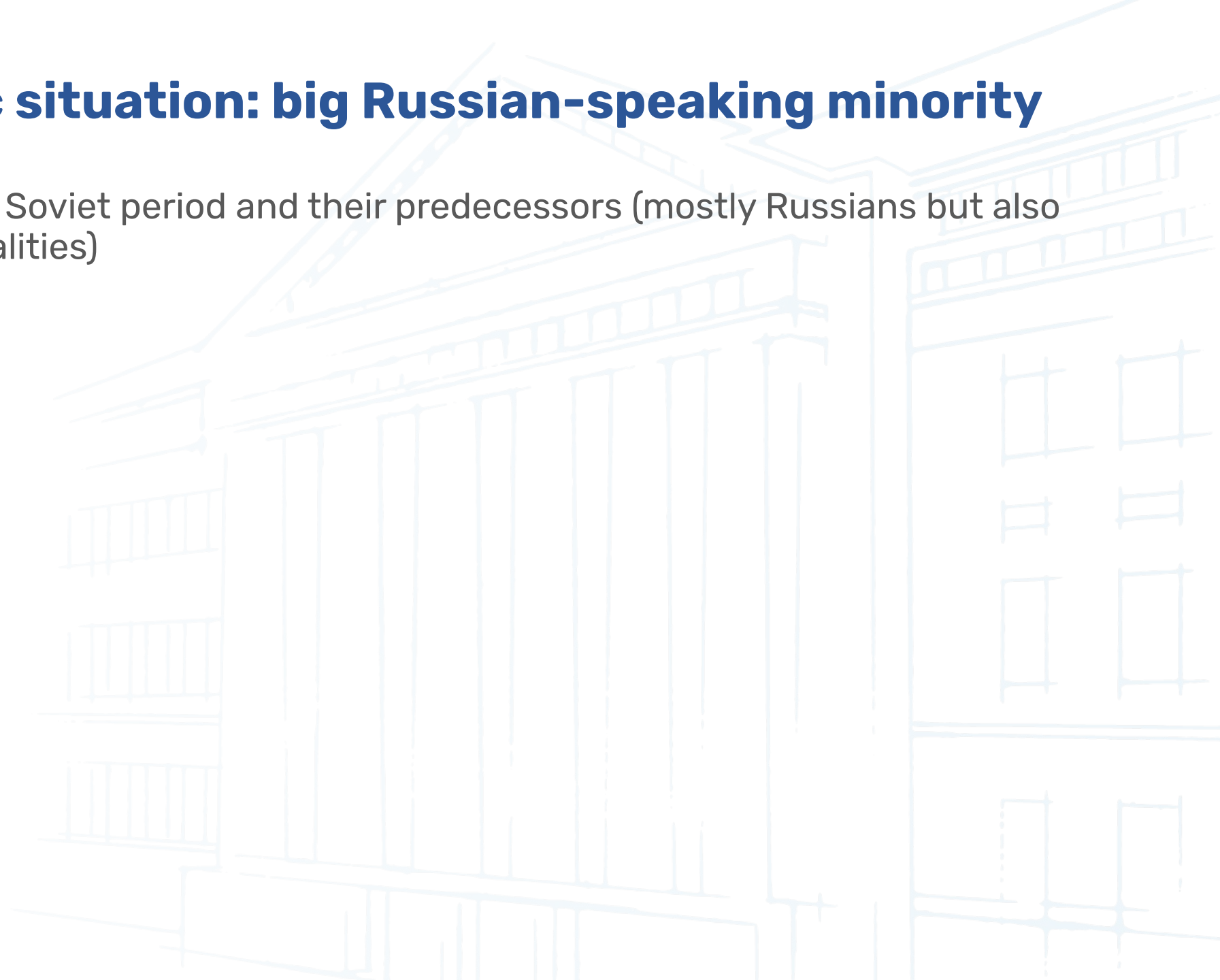




# **Russian minority: transition to Estonian school system**

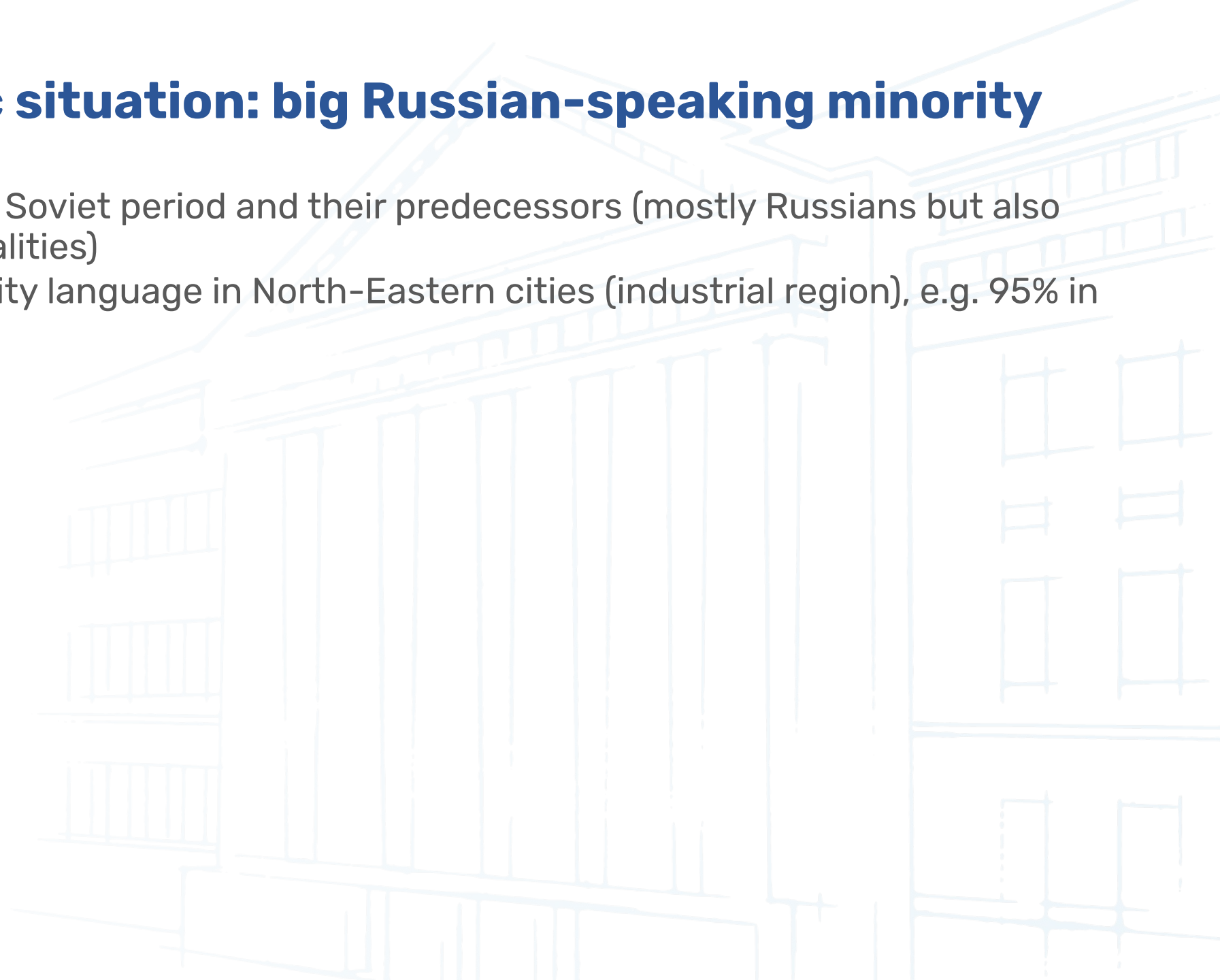
# **Sociolinguistic situation: big Russian-speaking minority**

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)



# **Sociolinguistic situation: big Russian-speaking minority**

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva



# **Sociolinguistic situation: big Russian-speaking minority**

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva
- two separate school systems herited from the Soviet period

# Sociolinguistic situation: big Russian-speaking minority

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva
- two separate school systems herited from the Soviet period
- teaching Estonian has been a challenge ( lack of motivation, lack of teachers) → young people with bad knowledge of Estonian → problems in job market

# Sociolinguistic situation: big Russian-speaking minority

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva
- two separate school systems herited from the Soviet period
- teaching Estonian has been a challenge ( lack of motivation, lack of teachers) → young people with bad knowledge of Estonian → problems in job market
- decision of transition to Estonian school (Estonian as a language of instruction) was made in 2022, the process started in 2024

# **Sociolinguistic situation: big Russian-speaking minority**

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva
- two separate school systems inherited from the Soviet period
- teaching Estonian has been a challenge ( lack of motivation, lack of teachers) → young people with bad knowledge of Estonian → problems in job market
- decision of transition to Estonian school (Estonian as a language of instruction) was made in 2022, the process started in 2024
- need for more teachers, teaching materials, teaching methods

# Sociolinguistic situation: big Russian-speaking minority

- migrants from the Soviet period and their predecessors (mostly Russians but also from other nationalities)
- Russian is a majority language in North-Eastern cities (industrial region), e.g. 95% in Narva
- two separate school systems inherited from the Soviet period
- teaching Estonian has been a challenge ( lack of motivation, lack of teachers) → young people with bad knowledge of Estonian → problems in job market
- decision of transition to Estonian school (Estonian as a language of instruction) was made in 2022, the process started in 2024
- need for more teachers, teaching materials, teaching methods
- need for re-thinking of Estonian school system in general (e.g. how minorities' mother tongue will be supported, teaching L1 and L2 speakers together)



# **Loss of dialects and the question of South Estonian language**

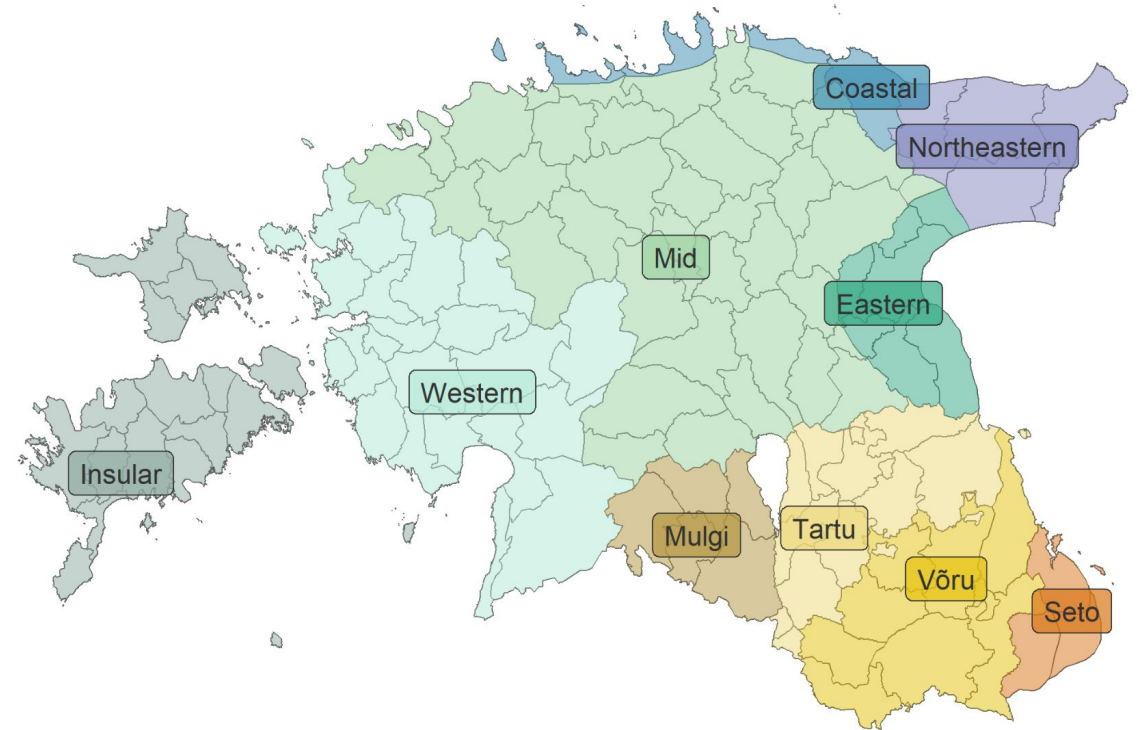
# Estonian dialects

Considerable linguistic variation in the traditional dialects.

Main differences between **North** and **South Estonian**, which origin from different tribal languages.

Coastal and Northeastern dialects have many common features with Finnish and other Eastern Finnic varieties.

Dialects have levelled; number of speakers is low.

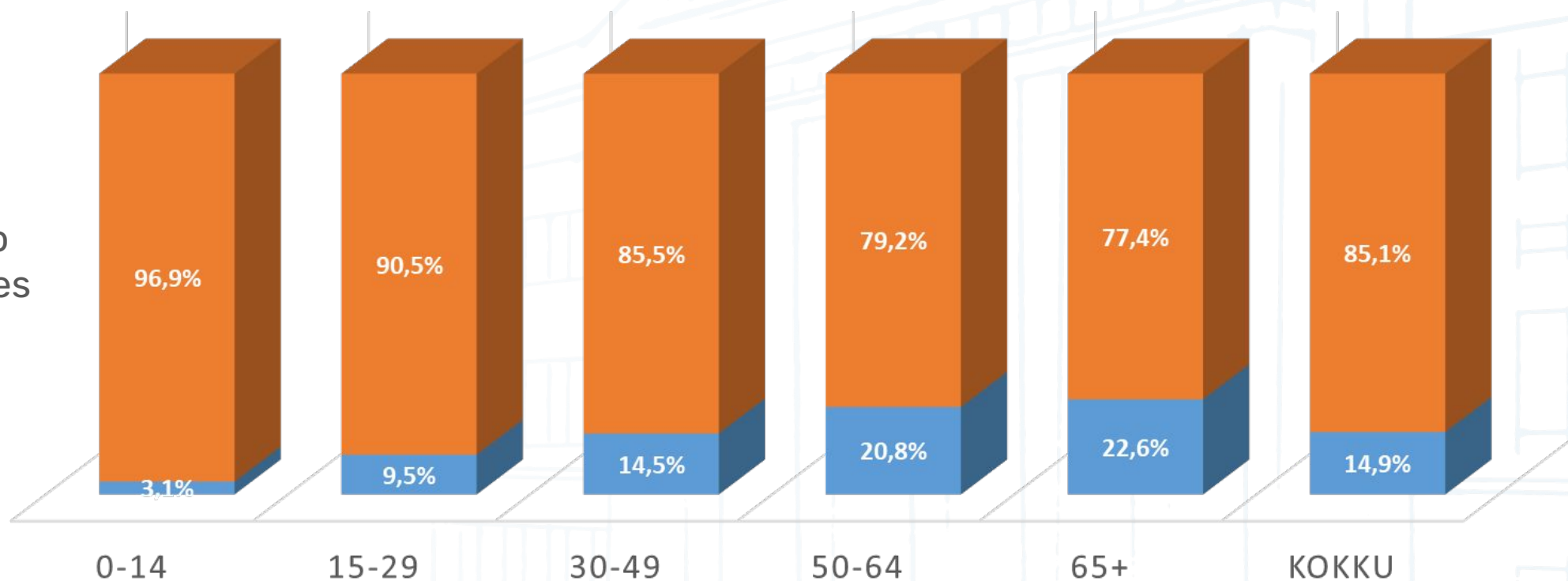


# Census 2021: Do you speak a dialect?

## MURDEKEELE OSKUS 2021

■ Oskab mõnda murret 2021 ■ Ei oska murret

Red: No  
Blue: yes



# South Estonian

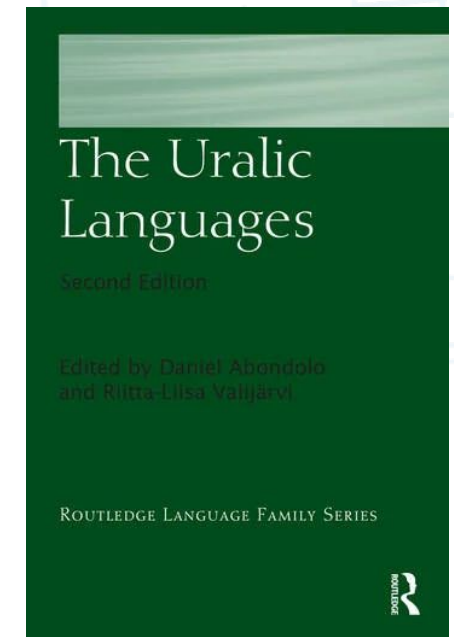
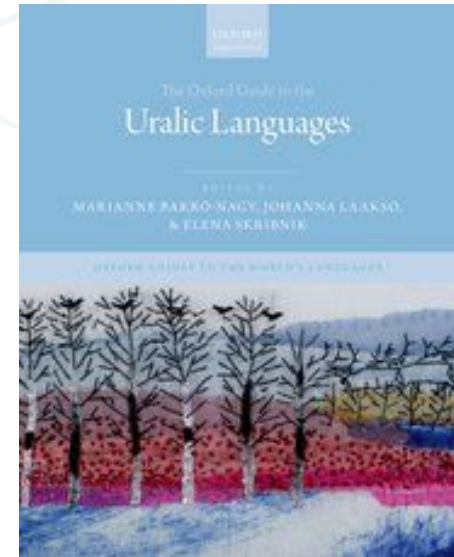
The number of speakers of SE varieties is more than **130,000** (according to Census 2021). However, their knowledge of SE is often poor.

SE is an **endangered language**:

- the shift to North/Standard Estonian started in the 1960s (Lindström et al. 2024)
- the intergenerational transmission of SE has stopped: the number of families that speak SE with their kids is less than 20-30

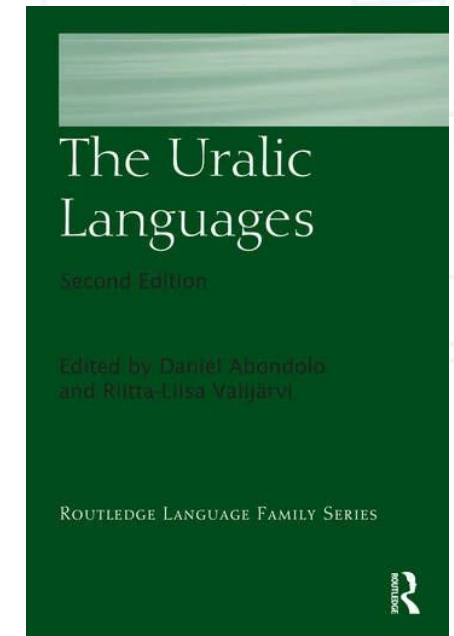
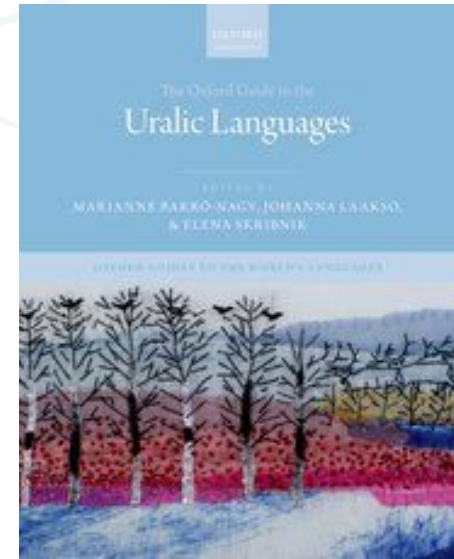
# South Estonian

- South Estonian has been recognized by linguists as a separate Finnic language, which diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2014, Prillop et al. 2020)
- International ISO code *Vro*; chapters in recent Uralic handbooks



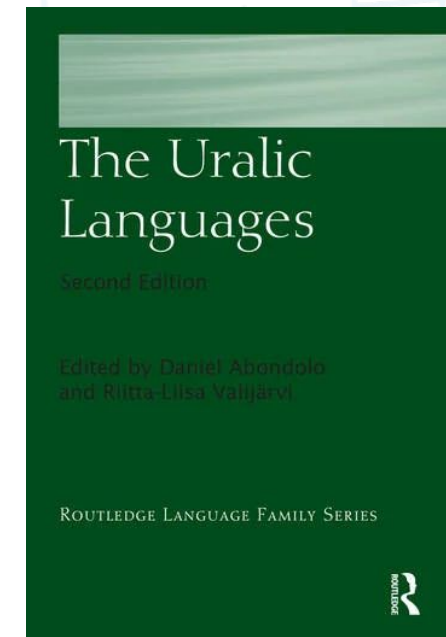
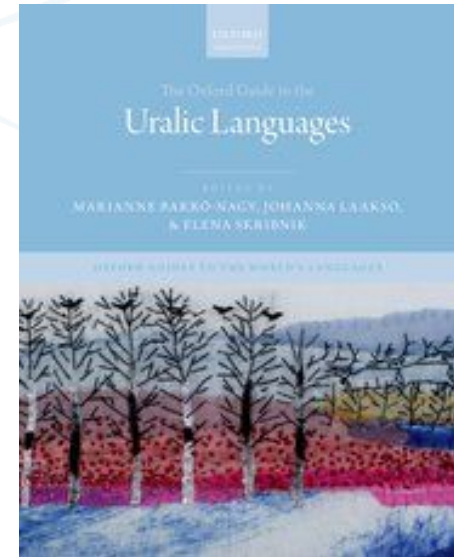
# South Estonian

- South Estonian has been recognized by linguists as a separate Finnic language, which diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2014, Prillop et al. 2020)
- International ISO code *Vro*; chapters in recent Uralic handbooks
- Also considered a language by its speakers (Pajusalu et al. 2000, etc.)



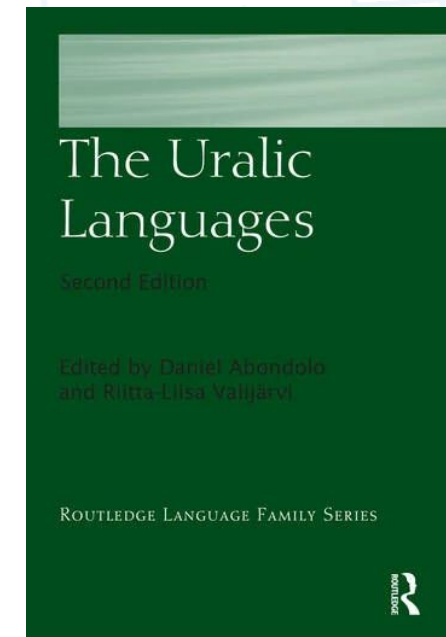
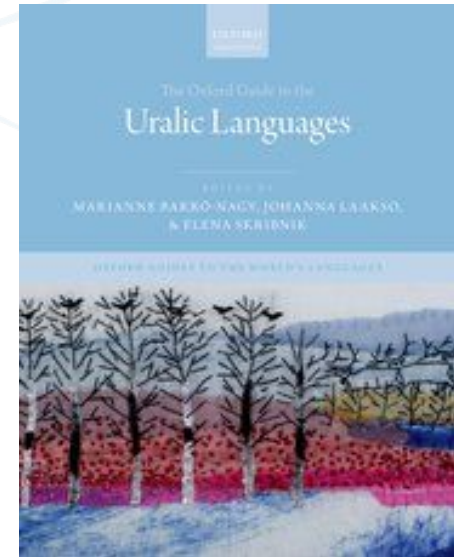
# South Estonian

- South Estonian has been recognized by linguists as a separate Finnic language, which diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2014, Prillop et al. 2020)
- International ISO code *Vro*; chapters in recent Uralic handbooks
- Also considered a language by its speakers (Pajusalu et al. 2000, etc.)
- Two written languages from 17th to 19th centuries: North Estonian (*Tallinna keel*) and South Estonian (*Tartu keel*)



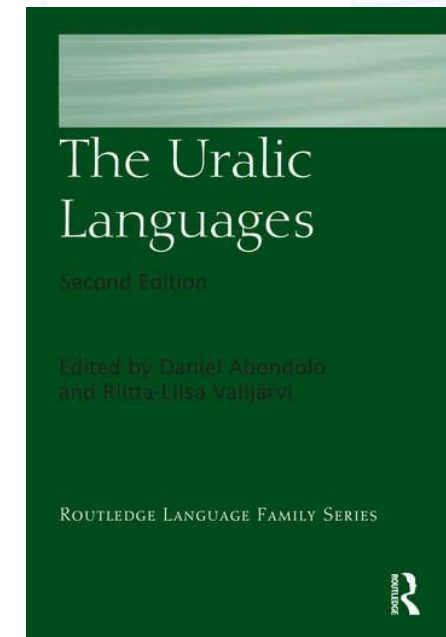
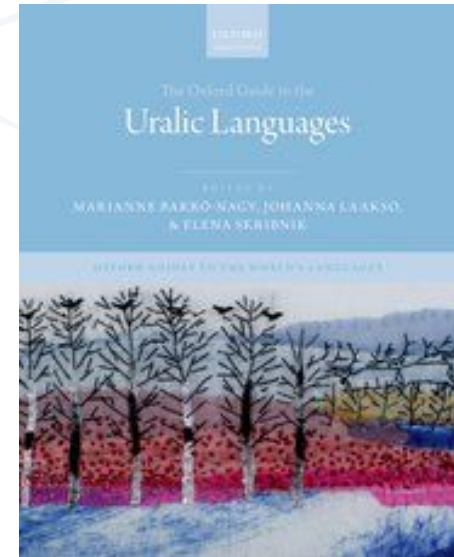
# South Estonian

- South Estonian has been recognized by linguists as a separate Finnic language, which diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2014, Prillop et al. 2020)
- International ISO code *Vro*; chapters in recent Uralic handbooks
- Also considered a language by its speakers (Pajusalu et al. 2000, etc.)
- Two written languages from 17th to 19th centuries: North Estonian (*Tallinna keel*) and South Estonian (*Tartu keel*)
- Traditional dialectology has not taken these facts into account: they still speak about *Estonian* dialects, which include South Estonian dialects (Võro, Seto, Mulgi, Tartu)



# South Estonian

- South Estonian has been recognized by linguists as a separate Finnic language, which diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2014, Prillop et al. 2020)
- International ISO code *Vro*; chapters in recent Uralic handbooks
- Also considered a language by its speakers (Pajusalu et al. 2000, etc.)
- Two written languages from 17th to 19th centuries: North Estonian (*Tallinna keel*) and South Estonian (*Tartu keel*)
- Traditional dialectology has not taken these facts into account: they still speak about *Estonian* dialects, which include South Estonian dialects (Võro, Seto, Mulgi, Tartu)
- In public discourse, traditional dialectology approach dominates



# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## **Problems:**

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media

# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## Problems:

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media
- the community is not homogeneous: different identities of Võro, Seto, Mulgi, Tartu speakers

# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## Problems:

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media
- the community is not homogeneous: different identities of Võro, Seto, Mulgi, Tartu speakers
- older SE speakers have experienced belittling of their language (esp. older generations) and do not consider their language valuable

# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## Problems:

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media
- the community is not homogeneous: different identities of Võro, Seto, Mulgi, Tartu speakers
- older SE speakers have experienced belittling of their language (esp. older generations) and do not consider their language valuable
- young generations do not speak SE (and do not fight for it)

# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## Problems:

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media
- the community is not homogeneous: different identities of Võro, Seto, Mulgi, Tartu speakers
- older SE speakers have experienced belittling of their language (esp. older generations) and do not consider their language valuable
- young generations do not speak SE (and do not fight for it)
- one state - one language ideology

# Discussion about the status of SE

No official recognition as a language (or: as a regional language, as a regional indigenous language) so far

Ongoing discussions about changes in the Estonian Language Act are not promising for SE

## Problems:

- no wider understanding about the SE: SE is far from Tallinn and the SE language is not seen or heard in public media
- the community is not homogeneous: different identities of Võro, Seto, Mulgi, Tartu speakers
- older SE speakers have experienced belittling of their language (esp. older generations) and do not consider their language valuable
- young generations do not speak SE (and do not fight for it)
- one state - one language ideology
- the Russian question

# Standard Language Ideology (SLI) in Estonia

The **belief that the standard language is the 'most correct'** or the 'best' form and that all other forms are 'incorrect' or less valid (Lippi-Green 2012: 67, Walsh 2021: 775).

The core of SLI is the belief in language correctness and in the one best language variety, leaving out and belittling all other varieties (Milroy 2001; Vogl 2012: 13).

SLI gives the standard language a greater level of **prestige** and a greater **legitimacy** over non-standard varieties.

SLI represented as 'common sense': when it comes to variation, it is considered a given that some forms are right and some are wrong; and the rules of grammar and senses of words exist **outside the speaker** (Milroy 2001).

Intertwined with **purism**: protection of a (prestigious) language from foreign influences and in some cases from dialects.

# SLI & purism in Estonia

1920s

- Quick development of standard language - new state needed a unified language form, which had to be taught at schools, used in government, legislation, army, etc.
- Great diversity in dialects.
- 'Cleaning' Estonian from German influences.

# SLI & purism in Estonia

## 1920s

- Quick development of standard language - new state needed a unified language form, which had to be taught at schools, used in government, legislation, army, etc.
- Great diversity in dialects.
- 'Cleaning' Estonian from German influences.

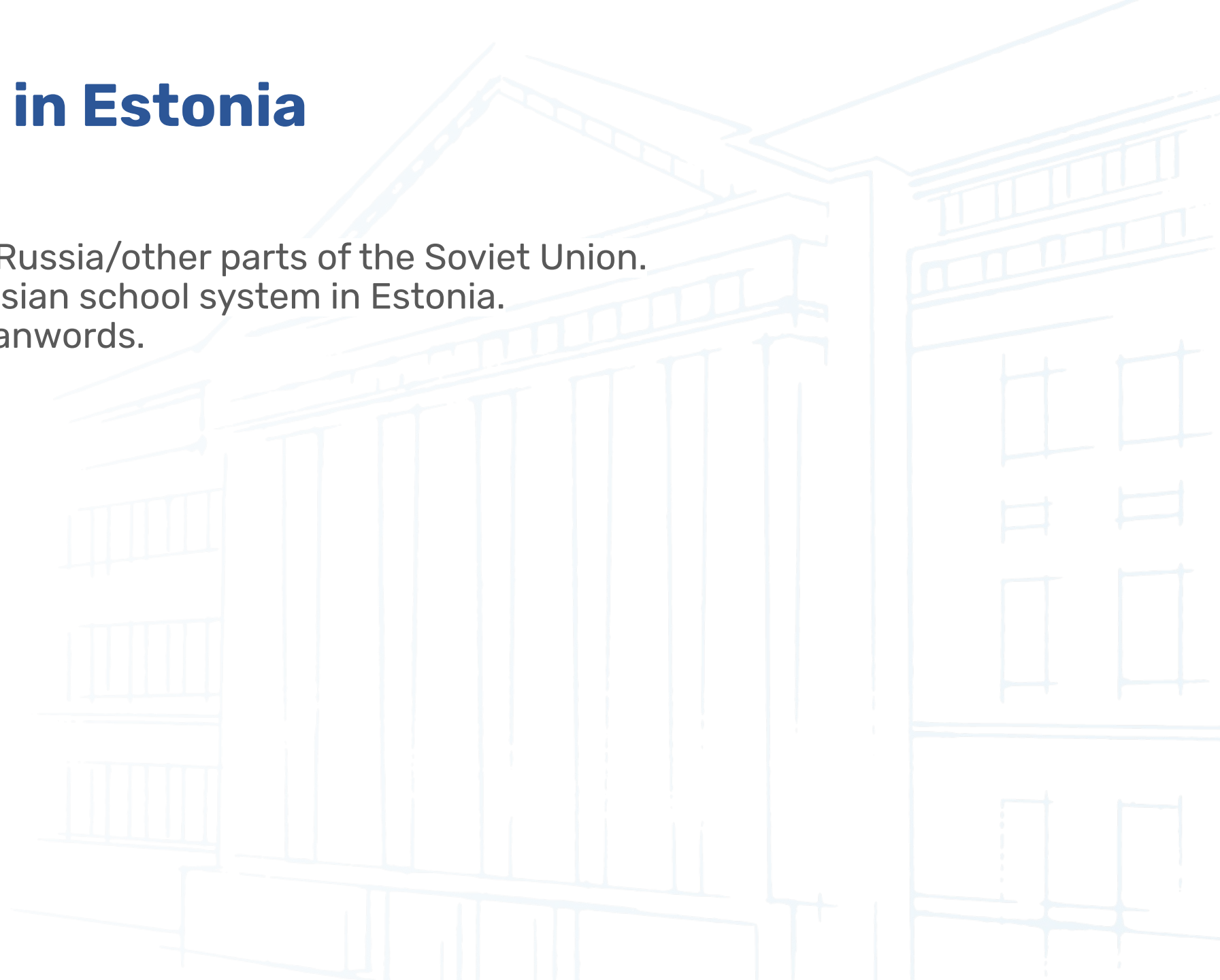
## 1930s

- The start of strict language planning (avoidance of parallel forms, 'one word - one meaning'), extreme systematicity (e.g., differentiating *ja* and *ning* 'and' etc.).
- Teaching 'correct' Estonian at schools was a priority.

# SLI & purism in Estonia

1940s-1950s

- Immigration from Russia/other parts of the Soviet Union.
- Establishing a Russian school system in Estonia.
- A lot of Russian loanwords.



# SLI & purism in Estonia

1940s-1950s

- Immigration from Russia/other parts of the Soviet Union.
- Establishing a Russian school system in Estonia.
- A lot of Russian loanwords.

1950s-1960s

- Conservative and strict language planning - no parallel forms, avoidance of polysemy.
- Dialect accents and errors were disapproved, especially in schools.

# SLI & purism in Estonia

1940s-1950s

- Immigration from Russia/other parts of the Soviet Union.
- Establishing a Russian school system in Estonia.
- A lot of Russian loanwords.

1950s-1960s

- Conservative and strict language planning - no parallel forms, avoidance of polysemy.
- Dialect accents and errors were disapproved, especially in schools.

1970s-1980s

- Slight liberalization of the standard language, e.g., more parallel forms allowed.
- “Pure” standard language as a means for resisting Russification.

# SLI & purism in Estonia

## 1940s-1950s

- Immigration from Russia/other parts of the Soviet Union.
- Establishing a Russian school system in Estonia.
- A lot of Russian loanwords.

## 1950s-1960s

- Conservative and strict language planning - no parallel forms, avoidance of polysemy.
- Dialect accents and errors were disapproved, especially in schools.

## 1970s-1980s

- Slight liberalization of the standard language, e.g., more parallel forms allowed.
- “Pure” standard language as a means for resisting Russification.

## 1980s-1990s

- Fight against Russian influences.
- Negative attitudes towards bilingualism (= obligatory Russian for other nationalities in the Soviet Union).

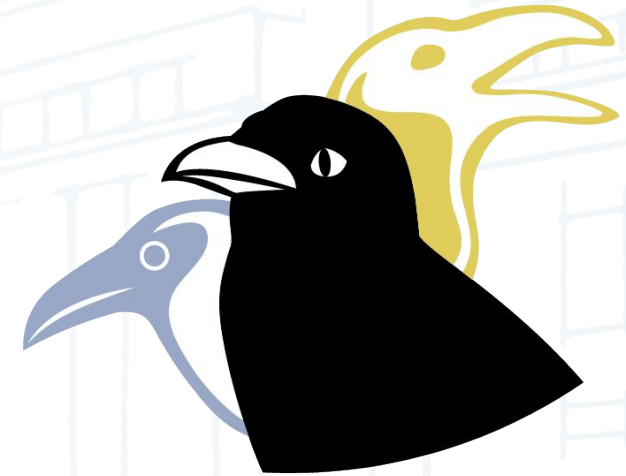
# SLI & purism in Estonia

From 1990s onwards:

- Less control over the language use, role of internet / social media, start of the 'destandardization' process.
- A lot of English loanwords.
- English influence is seen as a threat to Estonian.
- A desire for pure and beautiful Estonian (= standard Estonian) strengthened in the 2020s.
- Grassroot prescriptivism (e.g., in social media).

# VarEs: Morphosyntactic variation in Estonian

- 2024–2027
- Focus on **morphological and syntactic variation**
  - across different registers (e.g. newspapers, blogs, fora, edited and unedited texts)
  - across different modalities (spoken, written)
- Historical and dialectal background of variation patterns
- Spread and use of non-standard forms
- Attitudes towards linguistic variation
- Processing variation

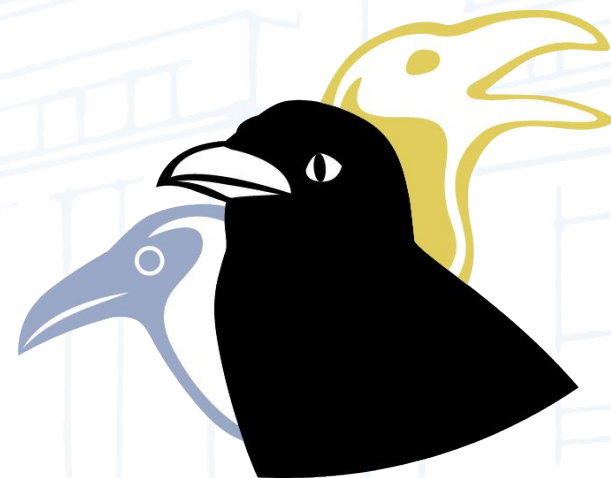


## VarEs

Morphosyntactic  
variation in Estonian

# VarEs: Morphosyntactic variation in Estonian

- 2024–2027
- Focus on **morphological and syntactic variation**
  - across different registers (e.g. newspapers, blogs, fora, edited and unedited texts)
  - across different modalities (spoken, written)
- Historical and dialectal background of variation patterns
- Spread and use of non-standard forms
- **Attitudes** towards linguistic variation
- **Processing** variation



## VarEs

Morphosyntactic  
variation in Estonian



# **Case studies**

# Study 1: what do Estonian speakers find annoying?

13. august 2023 · 🌐

Vähe sellest, et "et" käib liiga sageli kõnes lause algusesse, tundub "ikkagist" uus moesõna olevat. Eriti palju on seda kuulda telepurki sattunud noorte ja edukate suust. Iga kord kõlab see nagu keelekõrvakiil. Võeh! ja oeh! - ei muud. Aitäh, et kurtmise ära lugesite. 😊

👍 60

16. veebruar · 🌐

Eesti keeles on olemas kaks sõna - "tānu" ja "tõttu". Millegipärast on see viimane jäänud päris üheks sattunud. Miks? Kuidagi õõvastav on kuulda-lugeda, et ukrainlased on kaotanud sõjas nii palju mehi TĀNU ... Või teine näide: Eesti majandus on juba pikka aega languses TĀNU ... Mina kipun tänama ikka millegi hea eest ja seepärast eelpool mainitu häirib nii silma, kõrva kui ka mõistust.

1y Meeldib Vasta Jaga

25 kommentaari

1y 1h

Nööd (nõbud), brokoli (brokkoli), enne (ennem), kostab (kostub) jpm. Aga häirivaimad on siin juba kordi nimetatud dušš, kohv ja puder.

1y 2h

Ma olen ka eripedagoog ja mind ajab närvi, kui valesti kirjutavad ja räägivad inimesed, kes seda kindlasti teha ei tohiks. Näiteks eesti keele õpetaja ütleb taodelda, saatejuht raadios küsib, kuidas hindate presidendi kõne ja kirss tordil - eripedagoog ütleb, et kiidan lapse pere!

👍 2 💬 1 🔄 🚩

12h

Kui on üks sõna, mis mulle närvidele käib, siis on see "episood". Sarjadel on osad. Mitte episoodid.

👍 16 💬 13 🔄 🚩

11h

Tead seda ütlust, et haridus ei tähenda haritust? 😊 Mu meelest, see "kellegil" on täpselt see sõna millega haritust mõõta. Mulle juba öeldi, et ei tohi kõiki ühte patta panna, seega ma väidan, et 9/10-st "kellegil" kasutajatest on rumalad ja 1 tegi kogemata vea ja on tegelt norm. 😊

👍 💬 🔄 🚩

Suurim panustaja

Efektne, professionaalne, dušš, akutrell ... ilmselt on need juba mainitud. Siis lõpuks nagu -mata, -likkus, -gi/-ki. Sõnaalguline H on ka sassis, samuti b,d,g. Tegelikult tavalise inimese kirjakeelest kipubki jääma mulje, et vaesel talupojal ei olnud võimalust oma järglasi koolitada, ent järglasi sai palju.

1y Meeldib Vasta Jaga

4 👍

Anonüümne

28.01.2024 16:10:35

"Võru ja setu" keeled reostavad eesti keelt kümme korda rohkem, kui inglise või mõni muu keel ja neti släng! Võru keelsed koolid on isegi plaanis, mingeid metslasi kasvatama...

👤 Reageeri 🧑🏻👍👎😂😭 46

Laulupeo

28.01.2024 18:29:21

@Anonüümne Anonüümne 28.01.2024 17:18:54 Ja, kuidas on "teineteise" ja "üksteise" kasutamisega? Enne kasutati "teineteist", kui oli jutuks kahe inimese vahelisest tegevusest ja "üksteist" juba enama puhul. Aastaid olevat lubatud kasutada "üksteist" ka "teineteise" asemel. Ei pea õigeks...

Laulupeo: Eks see ole imelik jah kui öeldakse, et kaks inimest rääkisid üksteisega ja üksteist inimest rääkisid teineteisega.

👤 Reageeri 🧑🏻👍😂 5

Suurim panustaja

See pole nii otse kirjutamisega seotud ja minu elus ei muuda see midagi midagi ja see ei tohiks mulle pinda käia, aga käib 😊, kui inimesed ütlevad majasi, autosi jne. 😊

1y Meeldib Vasta Jaga

15 👍

# Study 1: what do Estonian speakers find annoying?

13. august 2023 · 🌐

It is bad enough that *et* ('that') so often starts a sentence – now *ikkagist* ('anyway') seems to be the new trendy word. You especially hear it from the young and successful people who end up on TV. Every time, it sounds like a linguistic slap in the face. Ugh! and oh! – nothing more. Thanks for reading my rant.

*Efektne* ('spectacular'), *professionaalne* ('professional'), *dušš* ('shower'), *akutrell* ('cordless drill')... I guess these have already been mentioned. Then there are the endings like *-mata*, *-likkus*, *-gi/-ki*. The initial *H* is also mixed up, as well as *b*, *d*, *g*. In fact, the written language of an average person gives the impression that the poor peasant never had the chance to educate his descendants – but he sure had plenty of them.

In Estonian, there are two words – *tänu* ('thanks to') and *tõttu* ('because of'). For some reason, the latter seems to have fallen out of favor. Why? It's somehow chilling to hear or read that Ukrainians have lost so many men in the war THANKS TO... Or another example: *The Estonian economy has been in decline for a long time THANKS TO...* I tend to give thanks only for something good, and that's why such uses as the ones above offend my eyes, ears, and mind alike.

**Anonüümne**

'Võru and Setu' dialects pollute the Estonian language ten times more than English or any other foreign language – or even internet slang! They're even planning Võru-language schools, to raise some kind of savages...

**Nõod** (*nõbud* 'cousins'), *brokoli* (*brokkoli* 'broccoli'), *enne* (*ennem* 'before'), *kostab* (*kostub* 'sounds'), and so on. But the most annoying ones are the often-mentioned *dušš* ('shower'), *kohv* ('coffee'), and *puder* ('porridge').

25 kommentaari

1y 1h 2h ... 12h ...

I'm also a special education teacher, and it drives me crazy when people who definitely shouldn't make mistakes write or speak incorrectly. For example, an Estonian language teacher says *taodelda* instead of *taotleda* ('to apply'), a radio host asks 'how do you evaluate the president's *kõne* ('speech')', and the cherry on top – a special ed teacher says 'I praise the child's *pere* ('family')'!

If there's one word that gets on my nerves, it's *episood* ('episode'). TV series have *osad* ('parts') not *episodes*.

16 13

I also don't know how to decline the word *kontsert* ('concert'), and I've very creatively rewritten sentences or replaced it with other words just to avoid having to double-check it again. Embarrassing, but that's just how it is. *Kohv* ('coffee') in the nominative case is often used wrong – it's especially ugly to see *kohvi* on a café sign.

1y Meeldib Vasta Jaga 9

You know the saying that being educated doesn't equal being knowledgeable? I think the word *kellegil* ('by someone') is exactly the kind of thing by which you can measure being knowledgeable. I've already been told not to lump everyone together, so I'll claim that 9 out of 10 *kellegil* users are stupid, and 1 just made a mistake and is actually normal.

11h ...

This isn't directly about writing, and it doesn't really change anything in my life, and it shouldn't bother me – but it does. When people say *majasi* ('houses'), *autosi* ('cars'), etc.

1y Meeldib Vasta Jaga 15

**Laulupeo**  
28.01.2024 18:29:21

And what about *teineteise* ('one another') and *üksteise* ('each other')? It used to be that *teineteise* was used for something between two people, and *üksteise* for more than two. For years now already, they've supposedly allowed *üksteise* to replace *teineteise*. I don't consider it right...

It is indeed strange when someone says that two people talked *üksteisega* ('to each other') and eleven people talked *teineteisega* ('to one another').

Reageeri 👍👎😄😅 5

# Study 1: what do Estonian speakers find annoying?

13. august 2023 · 🌐

It is bad enough that **et** ('that') so often starts a sentence – now **ikkagist** ('anyway') seems to be the new trendy word. You especially hear it from the young and successful people who end up on TV. Every time, it sounds like a linguistic slap in the face. Ugh! and oh! – nothing more. Thanks for reading my rant.

**Efektne** ('spectacular'), **professionaalne** ('professional'), **dušš** ('shower'), **akutrell** ('cordless drill')... I guess these have already been mentioned. Then there are the endings like **-mata**, **-likkus**, **-gi/-ki**. The initial **H** is also mixed up, as well as **b**, **d**, **g**. In fact, the written language of an average person gives the impression that the poor peasant never had the chance to educate his descendants – but he sure had plenty of them.

In Estonian, there are two words – **tänu** ('thanks to') and **tõttu** ('because of'). For some reason, the latter seems to have fallen out of favor. Why? It's somehow chilling to hear or read that Ukrainians have lost so many men in the war **THANKS TO**... Or another example: *The Estonian economy has been in decline for a long time THANKS TO*... I tend to give thanks only for something good, and that's why such uses as the ones above offend my eyes, ears, and mind alike.

**Nõod** (**nõbud** 'cousins'), **brokoli** (**brokkoli** 'broccoli'), **enne** (**ennem** 'before'), **kostab** (**kostub** 'sounds'), and so on. But the most annoying ones are the often-mentioned **dušš** ('shower'), **kohv** ('coffee'), and **puder** ('porridge').

25 kommentaari

Anonüümne

**'Võru and Setu' dialects** pollute the Estonian language ten times more than English or any other foreign language – or even internet slang! They're even planning Võru-language schools, to raise some kind of savages...

Reageeri 👍👎🤔😂 46

I'm also a special education teacher, and it drives me crazy when people who definitely shouldn't make mistakes write or speak incorrectly. For example, an Estonian language teacher says **taodelda** instead of **taotleda** ('to apply'), a radio host asks 'how do you evaluate the president's **kõne** ('speech')', and the cherry on top – a special ed teacher says 'I praise the child's **pere** ('family')'!

If there's one word that gets on my nerves, it's **episood** ('episode'). TV series have **osad** ('parts') not **episodes**.

👍 16 💬 13 🔄 🚩

I also don't know how to decline the word **kontsert** ('concert'), and I've very creatively rewritten sentences or replaced it with other words just to avoid having to double-check it again. Embarrassing, but that's just how it is. **Kohv** ('coffee') in the nominative case is often used wrong – it's especially ugly to see **kohvi** on a café sign.

1y Meeldib Vasta Jaga

9 👍

You know the saying that being educated doesn't equal being knowledgeable? I think the word **kellegil** ('by someone') is exactly the kind of thing by which you can measure being knowledgeable. I've already been told not to lump everyone together, so I'll claim that 9 out of 10 **kellegil** users are stupid, and 1 just made a mistake and is actually normal.

This isn't directly about writing, and it doesn't really change anything in my life, and it shouldn't bother me – but it does. When people say **majasi** ('houses'), **autosi** ('cars'), etc.

1y Meeldib Vasta Jaga

15 👍

Laulupeo

28.01.2024 18:29:21

And what about **teineteise** ('one another') and **üksteise** ('each other')? It used to be that **teineteise** was used for something between two people, and **üksteise** for more than two. For years now already, they've supposedly allowed **üksteise** to replace **teineteise**. I don't consider it right...

It is indeed strange when someone says that two people talked **üksteisega** ('to each other') and eleven people talked **teineteisega** ('to one another').

Reageeri 👍😂 5

# Study 1: what do Estonian speakers find annoying?

- An online survey conducted in February–March 2025 to
  - find out what types of language use people consciously notice the most and how much such language use bothers them;
  - collect input for an experiment aimed at studying **the processing of variable linguistic phenomena** during text reading and **the language attitudes** associated with these phenomena.
- Only adult L1 Estonian speakers.
- The survey link was shared mostly in various social media channels.

# Study 1: task

Palun nimeta kuni 10 eesti keele nähtust, mida emakeelsete eesti keele kõnelejate suulises või kirjalikus keelekasutuses sageli märkad ja tähele paned. Tähelepanekud võivad hõlmata nii sõnakasutust, tähendusi, õigekirja, käänamist-pööramist, lauseehitust kui ka hääldust. Hinda ka seda, kas ja kuivõrd kirjeldatud keelekasutus sind häirib.

Järgmist nähtust saad kirjeldada pärast seda, kui oled eelmise ära hinnanud. Nupu "Järgmine" vajutamine lõpetab loetelu ning suunab edasi lehele, kus saad soovi korral jätta lisakommentaare ja täpsustusi.

**Nähtus**

**Kus ja kelle keelekasutuses oled seda märganud?**

1.

Hinda, kui palju 1. kirjeldatud nähtus sind häirib.

**1 - ei häiri üldse**

**2**

**3**

**4**

**5 - häirib väga**

☐☐☐☐☐

# Study 1: task

Please name up to 10 features of Estonian language that you often notice in the spoken or written language use of native Estonian speakers. Your observations may include aspects such as word choice, meanings, spellings, inflection or conjugation, sentence structure, or pronunciation. Also assess whether, and to what extent, the described language use bothers you.

You can describe the next feature after you have evaluated the previous one. Clicking the 'Next' button will end the list and take you to a page where you can, if you wish, leave additional comments and clarifications.

**Observation**

**Where and in whose language use have you noticed it?**

1.

Evaluate how much the 1st described phenomenon bothers you.

**1 – not at all**

☐

**2**

☐

**3**

☐

**4**

☐

**5 – very much**

☐

# Study 1: minimal background information

- Additional information collected about the respondents:
  - gender
  - age
  - highest educational level obtained
  - whether their work involves working with texts in Estonian.

# Study 1: expectations

We expected that in a listing task, people will name more phenomena which

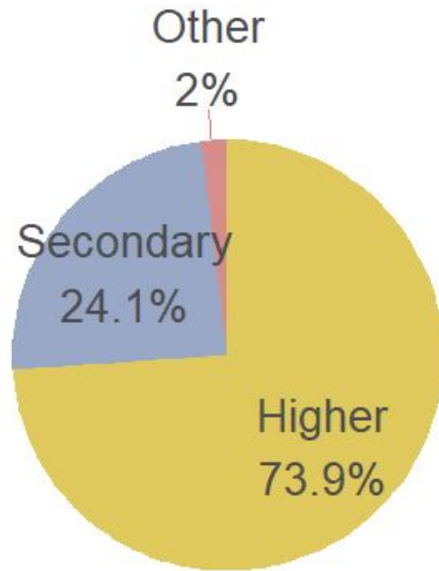
- are related to the standard language norms and are thus discussed in language planning sources and native language instruction;
- are perceived as strongly bothersome.

The ideology of the standard language inclines people toward an abstract, idealized, and homogenous notion of language (Lippi-Green 2012: 67), where the often artificial norms of the standard variety tend to be regarded as common sense or as something timeless and universally valid (Garrett 2010: 7).

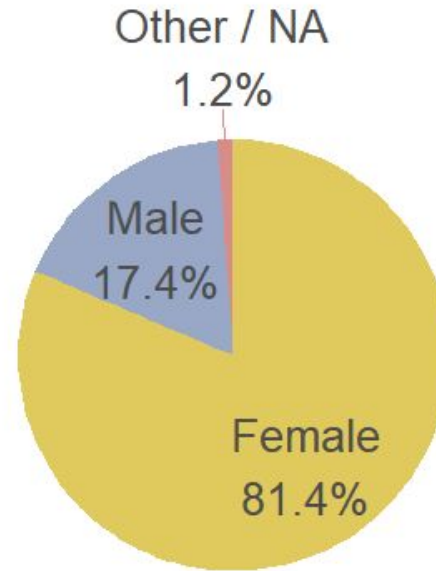
Linguistic variation is seen as a problem (Davies 2000: 122).

# Study 1: respondents

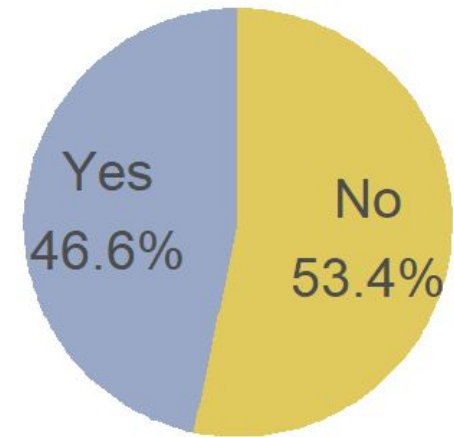
## Education



## Gender

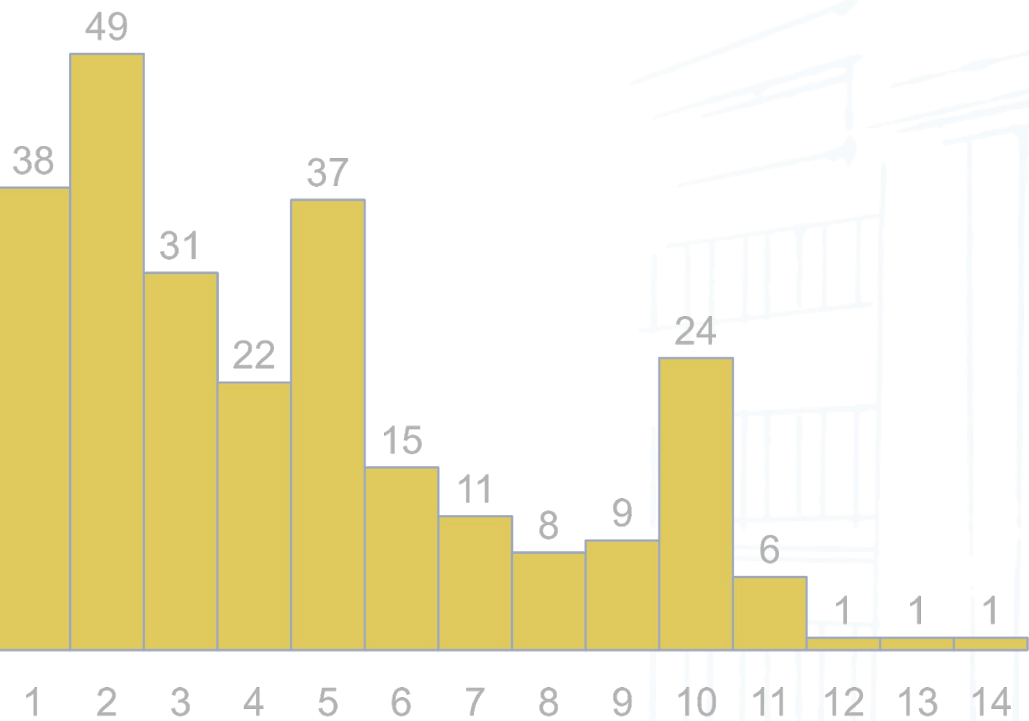


## Work with language

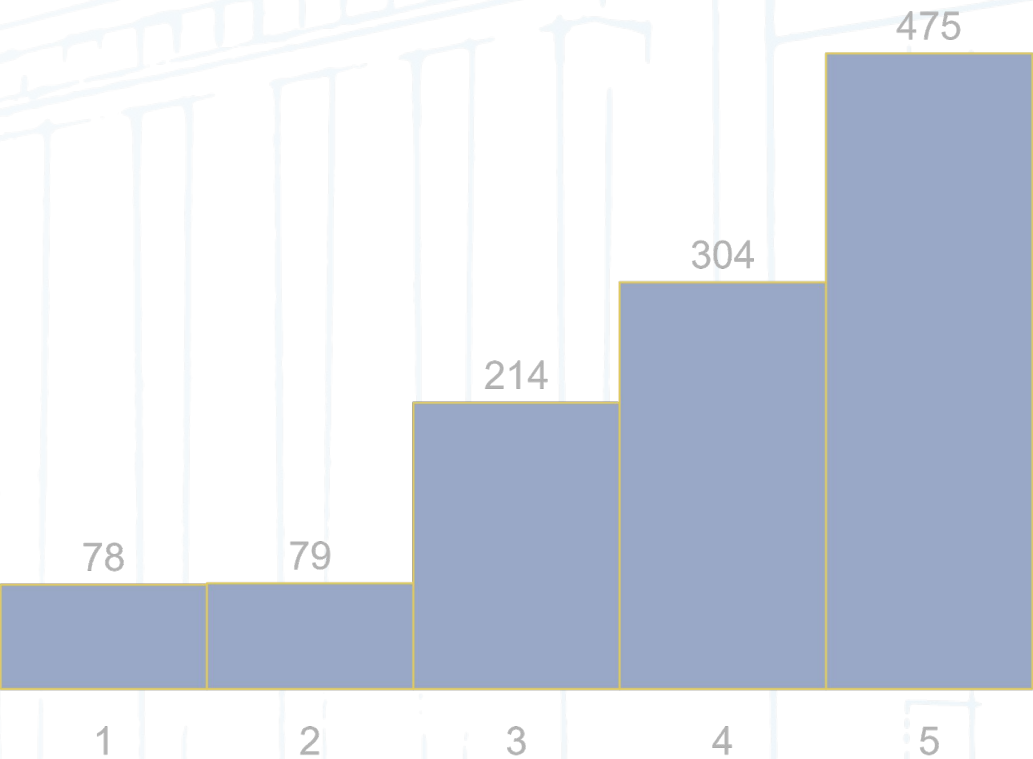


# Study 1: 1159 responses

Number of listed phenomena



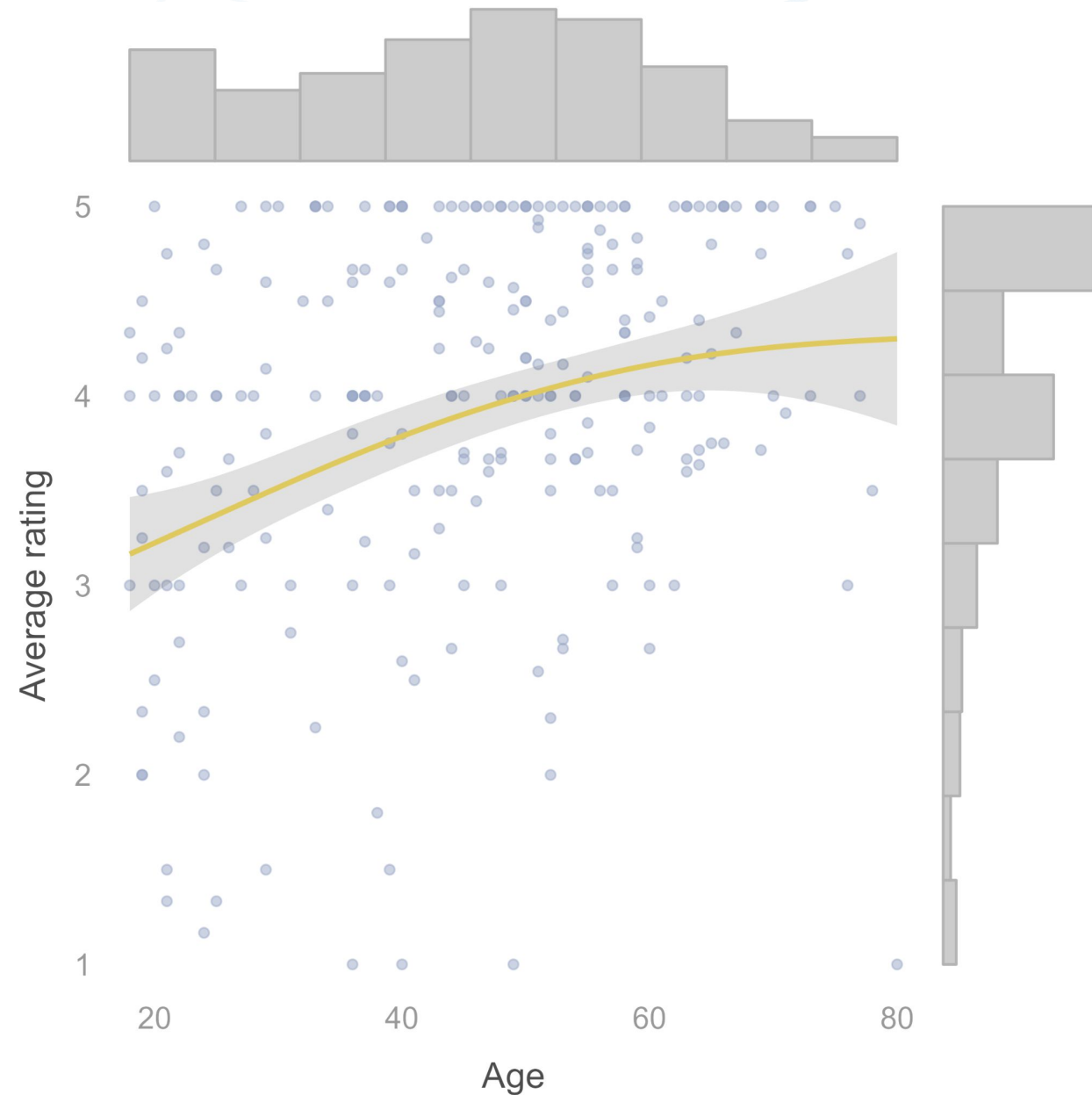
Ratings



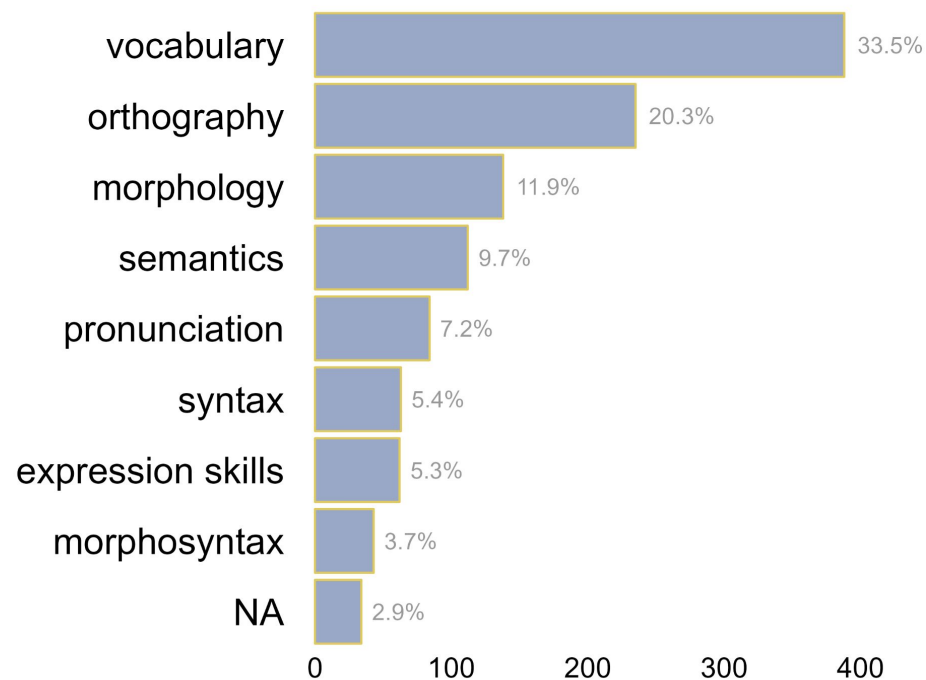
# Study 1: 1159 responses

The average ratings of the respondents did not correlate with their gender, education or work.

However, there is a correlation between the respondent's **age** and their **average rating**: the older the respondent, the more bothersome they found the phenomena they listed (GAM: edf = 1.965,  $F = 14.2$ ,  $p < .001$ ).



# Study 1: categories



## **Vocabulary:**

foreign words and expressions (33%);  
non-standard word forms and expressions (23.2%);  
discourse particles and markers (11.9%);  
raw translations (10.3%)

## **Orthography:**

punctuation (23.4%);  
spelling mistakes (16.2%);  
phonetic spelling (12.8%);  
compounding and separation (12.3%);  
capital letters (11.9%)

## **Morphology:**

declension type alternation (55.1%);  
the position of the *gi-/ki-*clitic (16.7%)

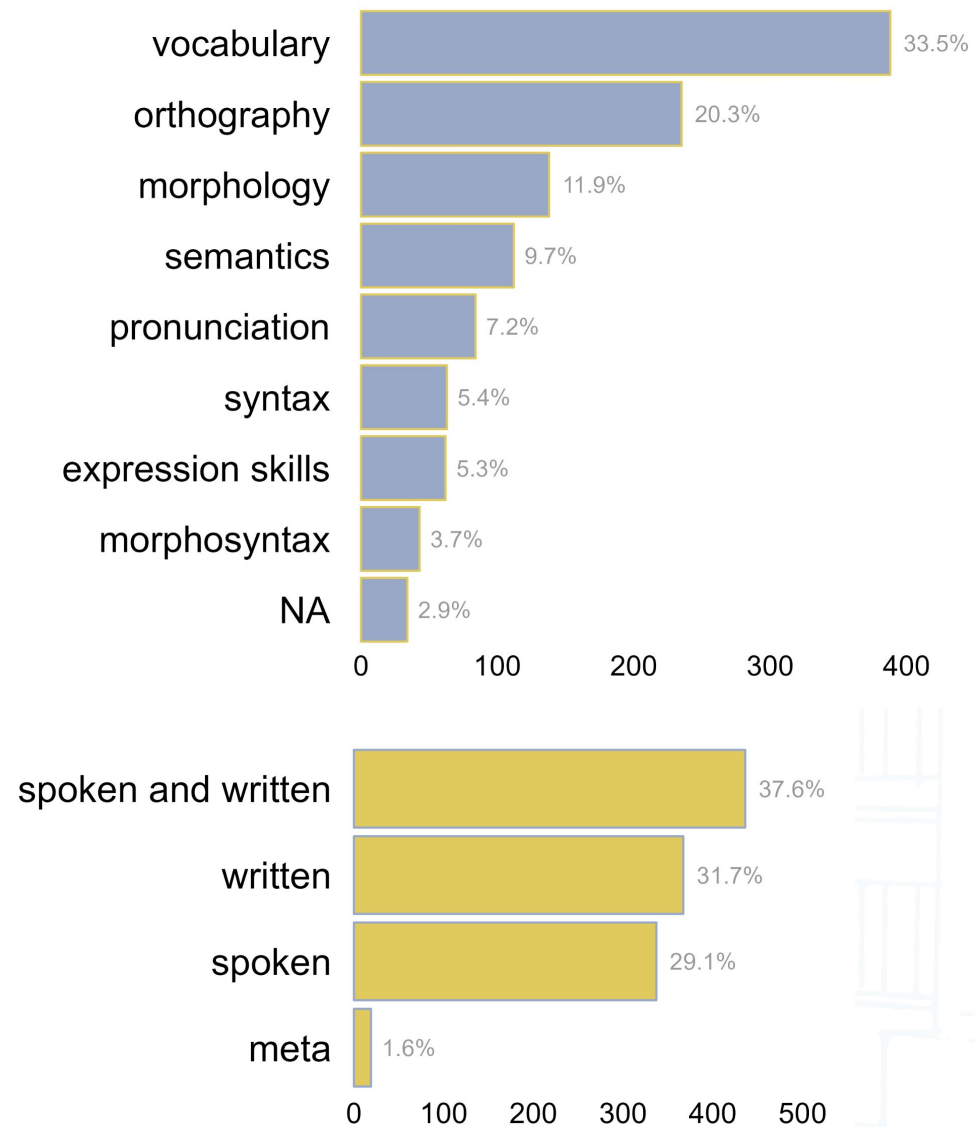
## **Semantics:**

non-standard meanings (58%);

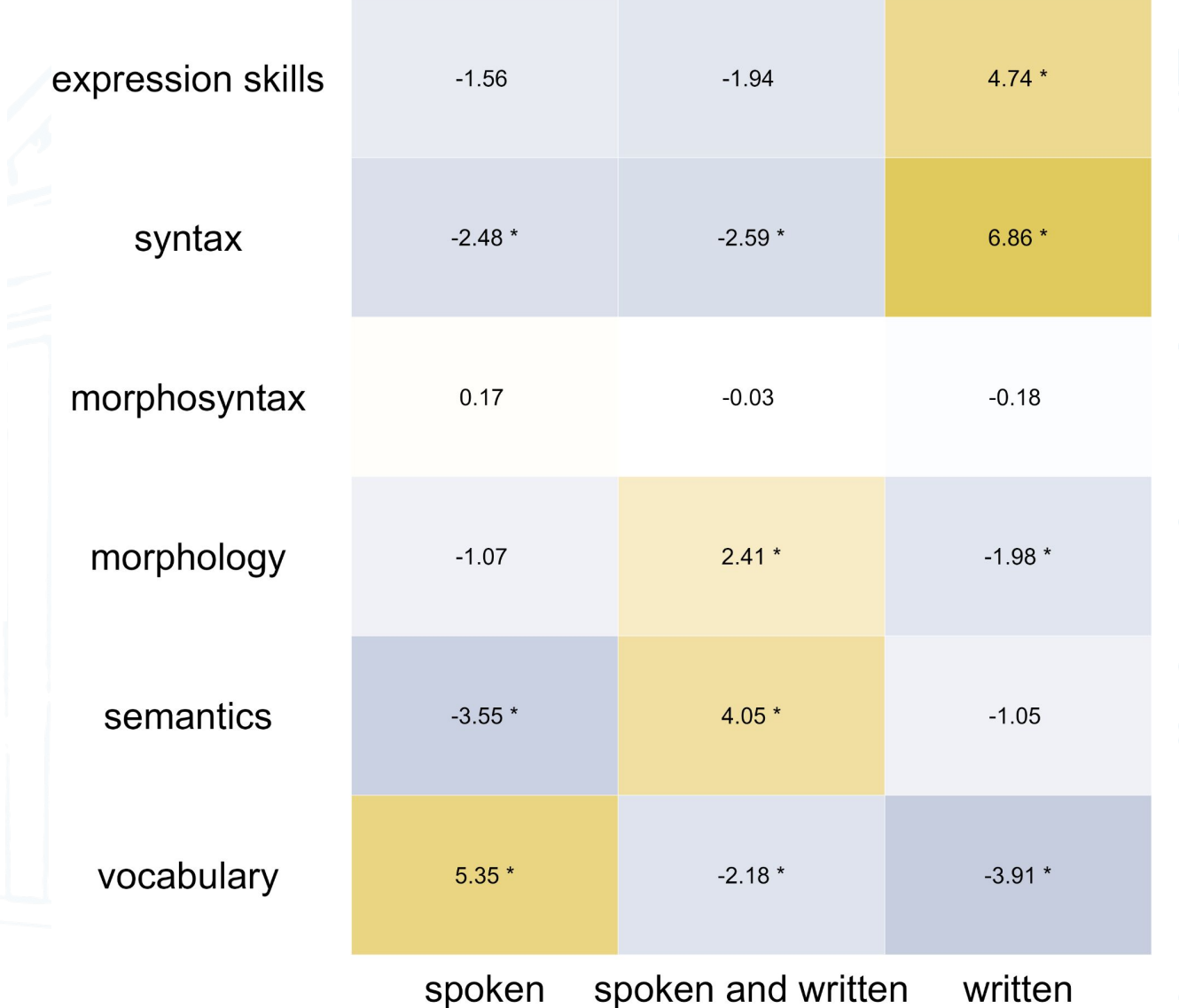
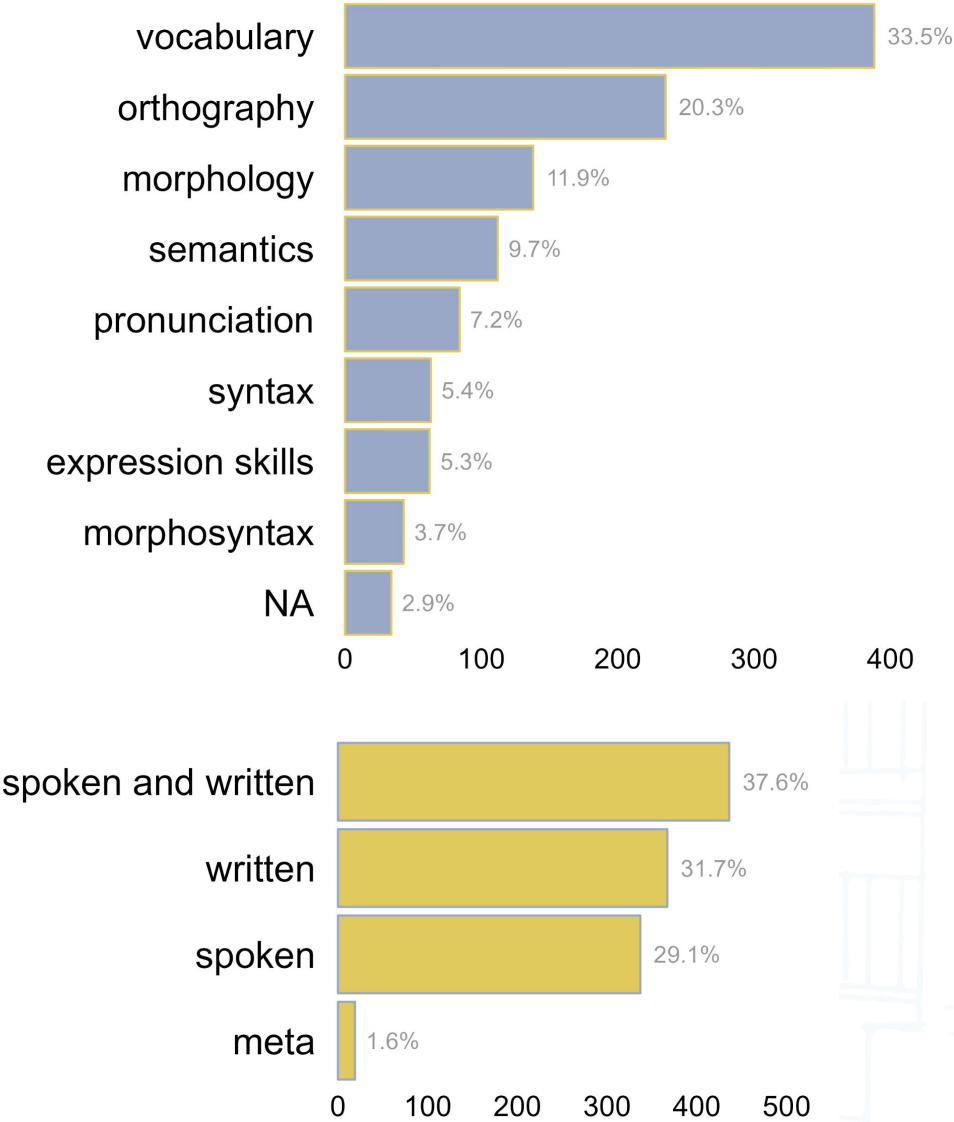
## **Pronunciation:**

phonological reduction (17.9%);  
word stress (11.9%)

# Study 1: categories



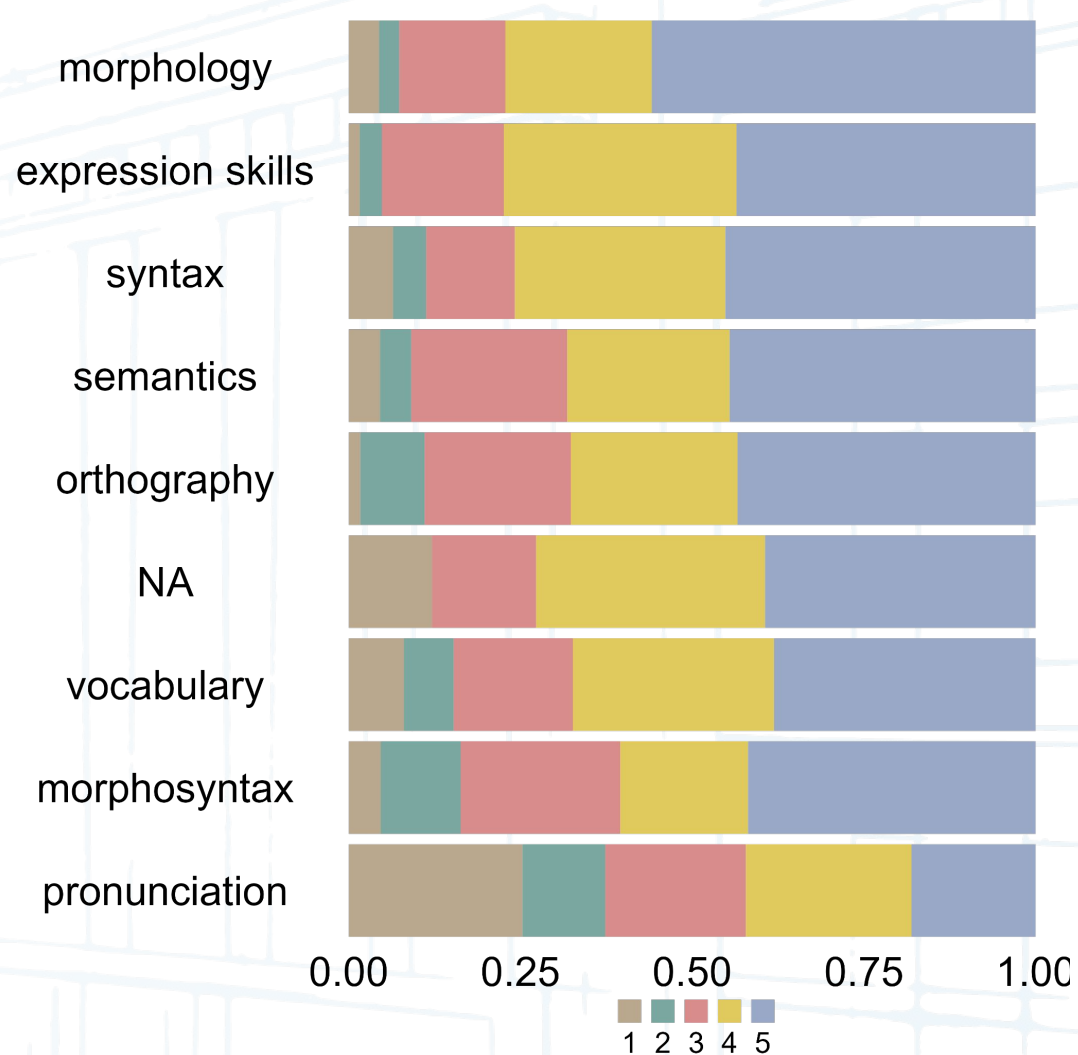
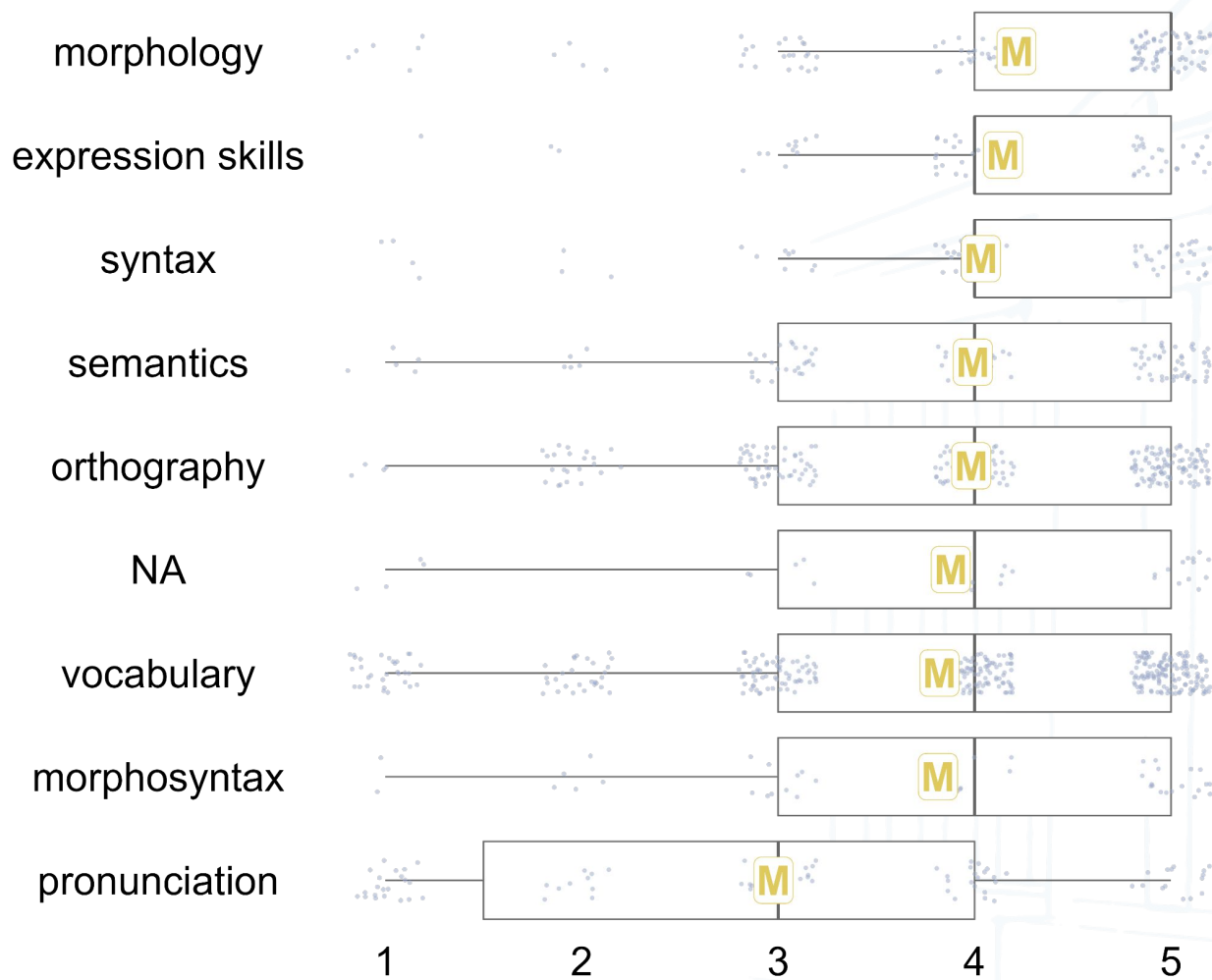
# Study 1: categories



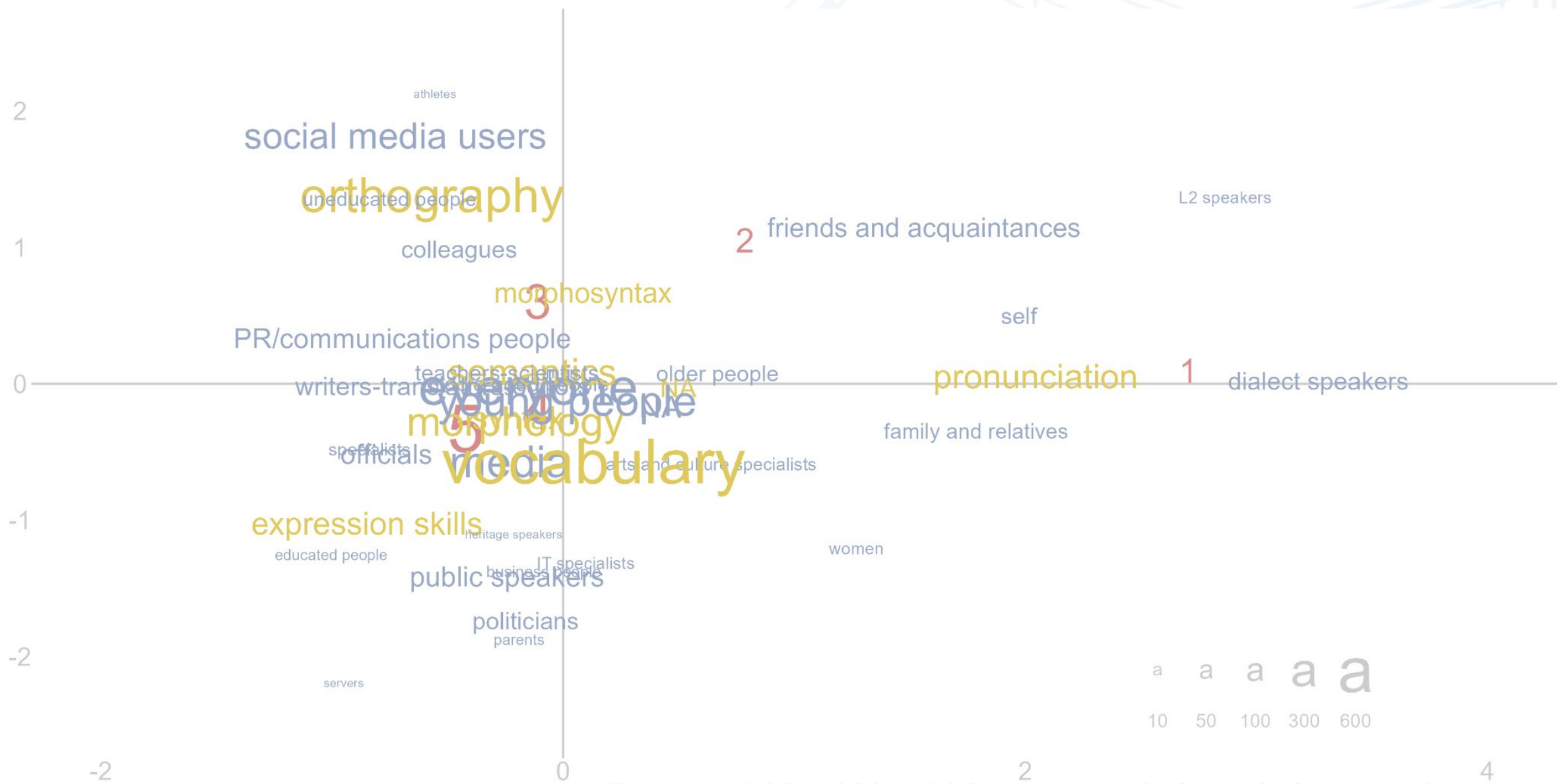
Pearson's standardised residuals

# Study 1: ratings

Distribution of ratings



# Study 1: MCA



# Summary of study 1

Based on the 1159 responses from 253 respondents,

- as expected, most of the phenomena listed were perceived as **strongly bothersome**;
- the rating **correlated with age**: the older the respondent, the more bothersome the listed phenomenon;
- the phenomena which were listed the most, represented categories which are practiced the most at school and/or which are the most salient / easily perceived:
  - **vocabulary**: the use and adaptation of foreign words and the use of non-standard forms and expressions
  - **orthography**
- phenomena related to **pronunciation** were found the least bothersome. However, the respondents are bothered by the expression of pronunciation in orthography or morphology.

# Study 2: Does morphological variation cause problems in reading?

- An eyetracking experiment to examine whether **morphologically varying forms** are read differently depending on
  - their normative status
    - 1) standard vs. non-standard variation  
*imet* ~ *ime* ('miracle' sg.par)
    - 2) both forms standard (*parallel forms*)  
*vaidlusi* ~ *vaidluseid* ('argument' pl.par)
  - their frequency
  - personality type of the participant (The Big Five Inventory)

Annika Kängsepp



Külli Habicht

# Study 2: Does morphological variation cause problems in reading?

- An eyetracking experiment to examine whether **morphologically varying forms** are read differently depending on
  - their normative status
    - 1) standard vs. non-standard variation  
*imet* ~ *ime* ('miracle' sg.par)
    - 2) both forms standard (*parallel forms*)  
*vaidlusi* ~ *vaidluseid* ('argument' pl.par)
  - their frequency
  - personality type of the participant (The Big Five Inventory)
- **Pilot study** conducted in September 2025:
  - 10 participants
  - 6 female, 4 male
  - age 22–36 years (mean 29.6)
  - 8 participants had a background in linguistics

Annika Kängsepp



Külli Habicht

## Study 2: Study design

Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.

60 standard-nonstandard

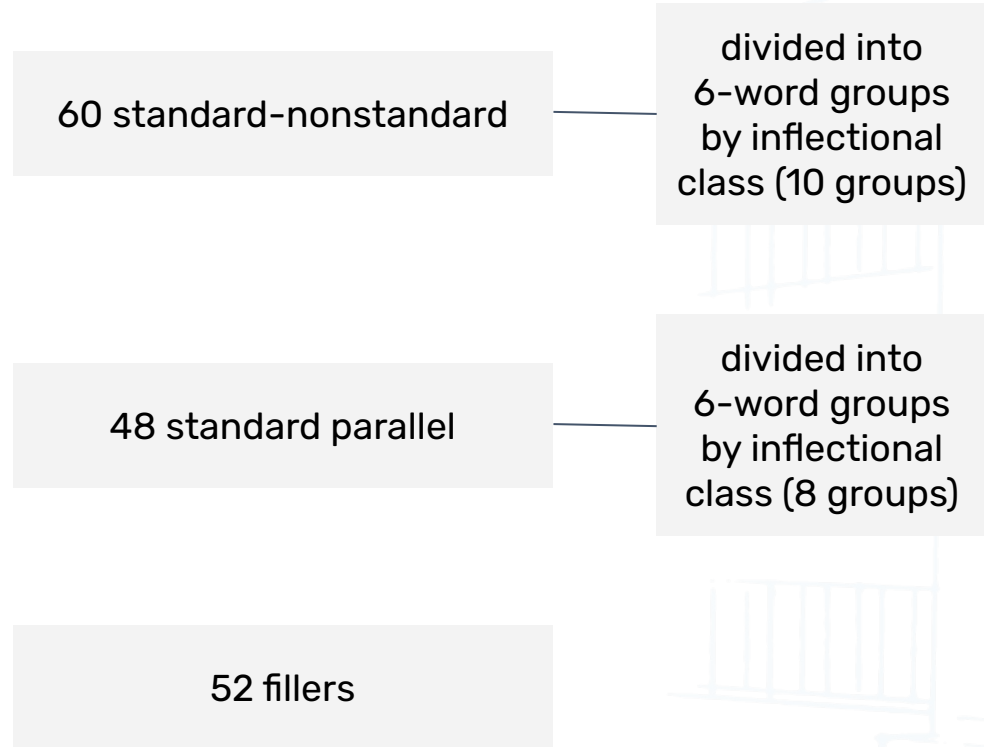
48 standard parallel

52 fillers



## Study 2: Study design

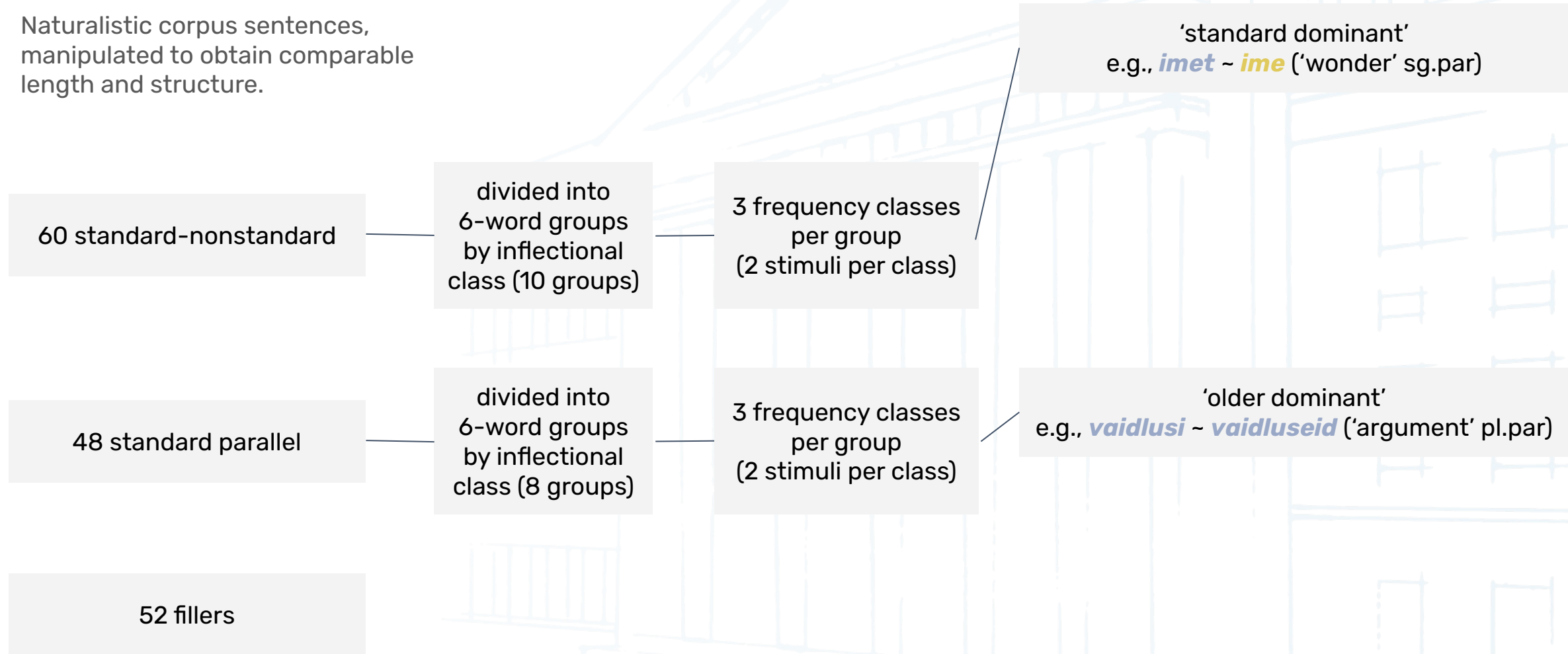
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



## Study 2: Study design

Based on frequencies from  
the Estonian National Corpus (2023),  
3.8 billion tokens

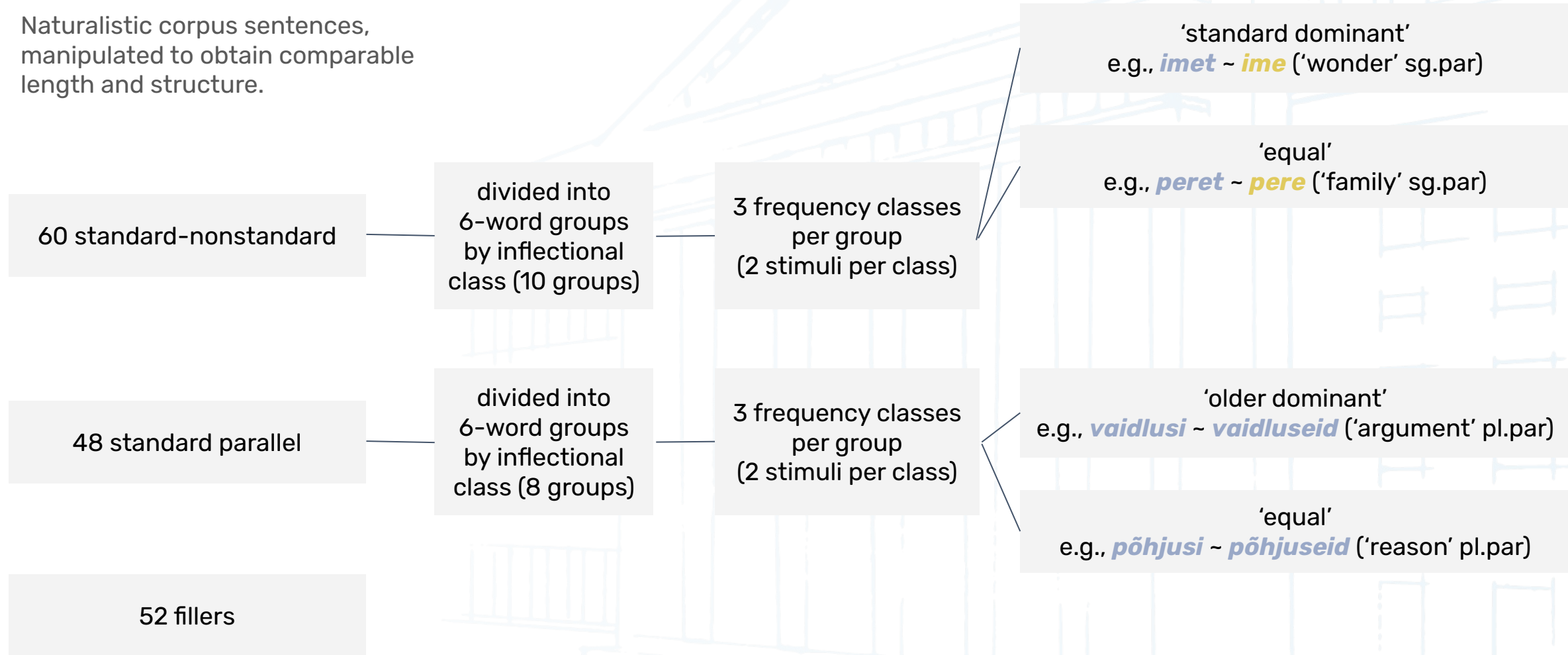
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



## Study 2: Study design

Based on frequencies from  
the Estonian National Corpus (2023),  
3.8 billion tokens

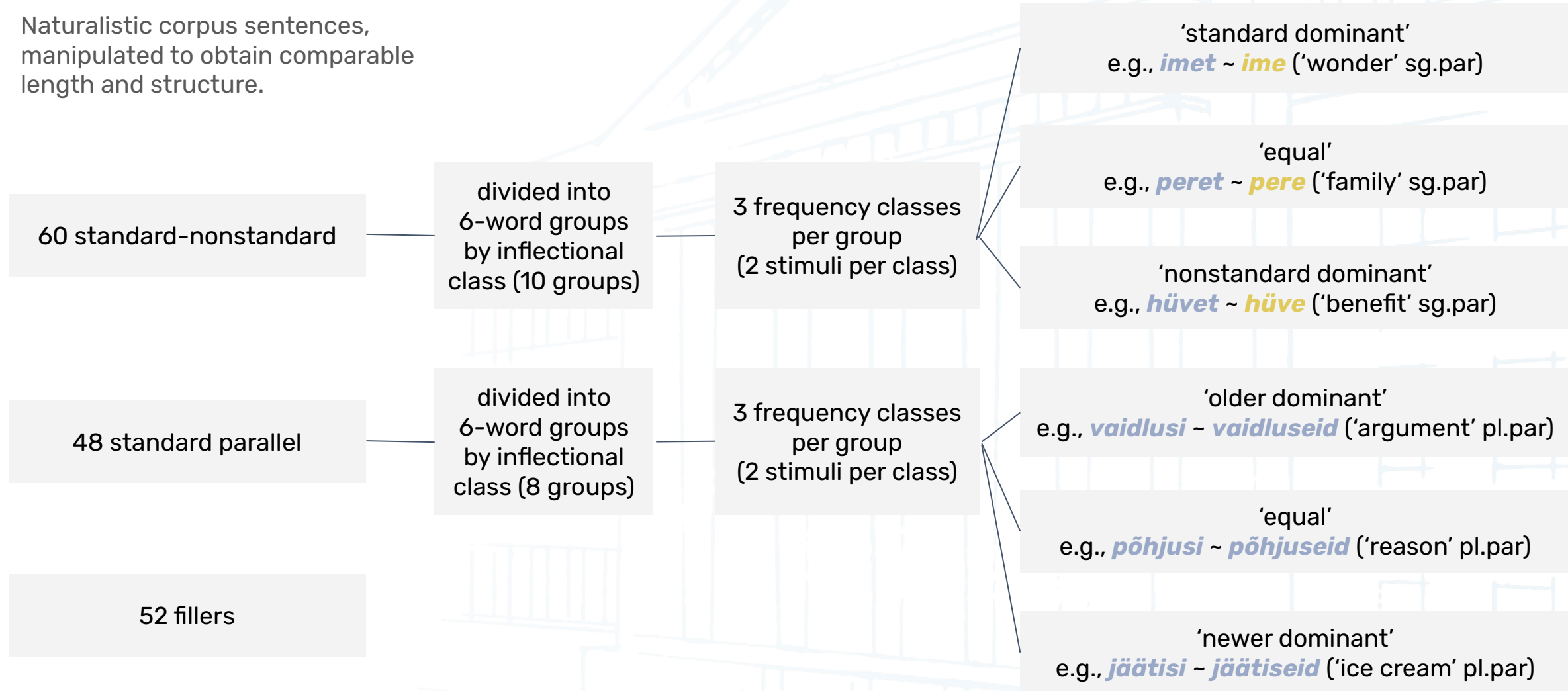
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



## Study 2: Study design

Based on frequencies from  
the Estonian National Corpus (2023),  
3.8 billion tokens

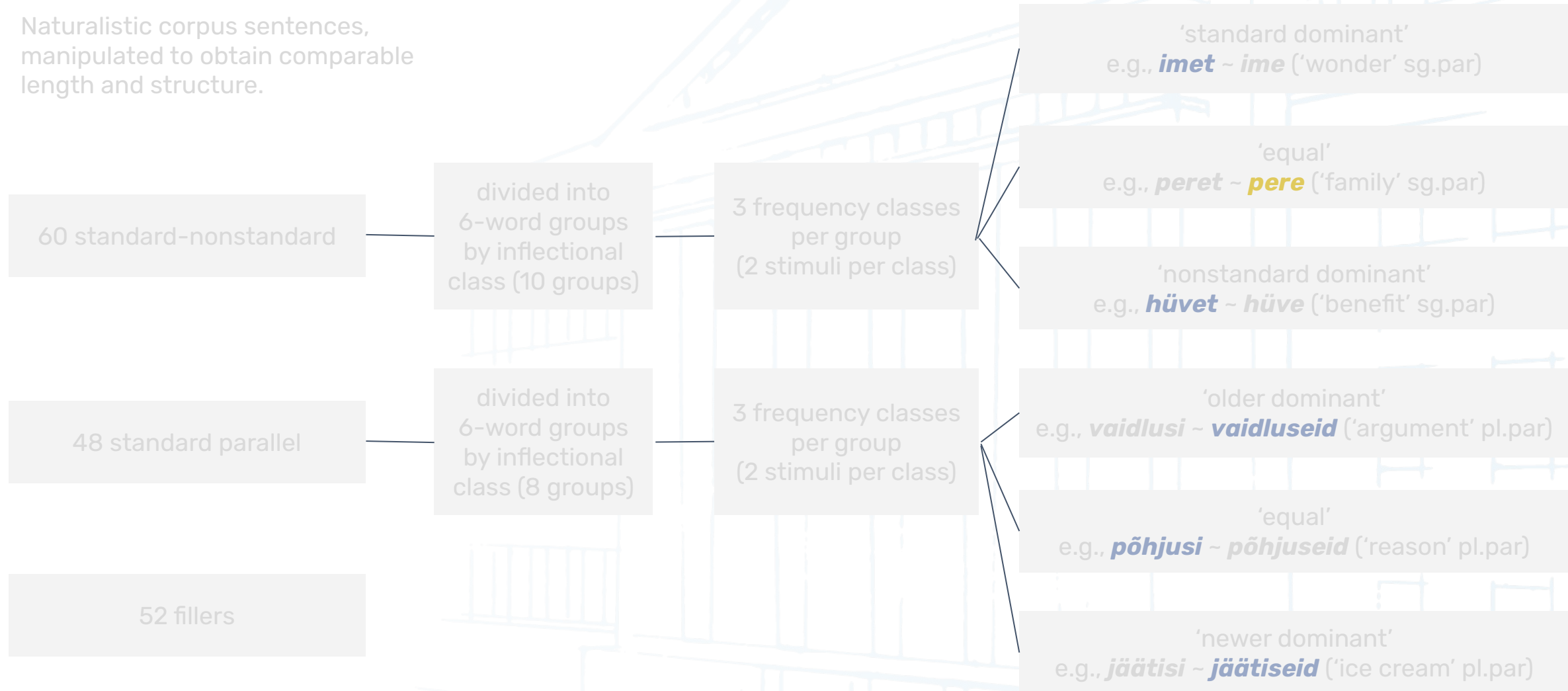
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



## Study 2: Study design

Based on frequencies from  
the Estonian National Corpus (2023),  
3.8 billion tokens

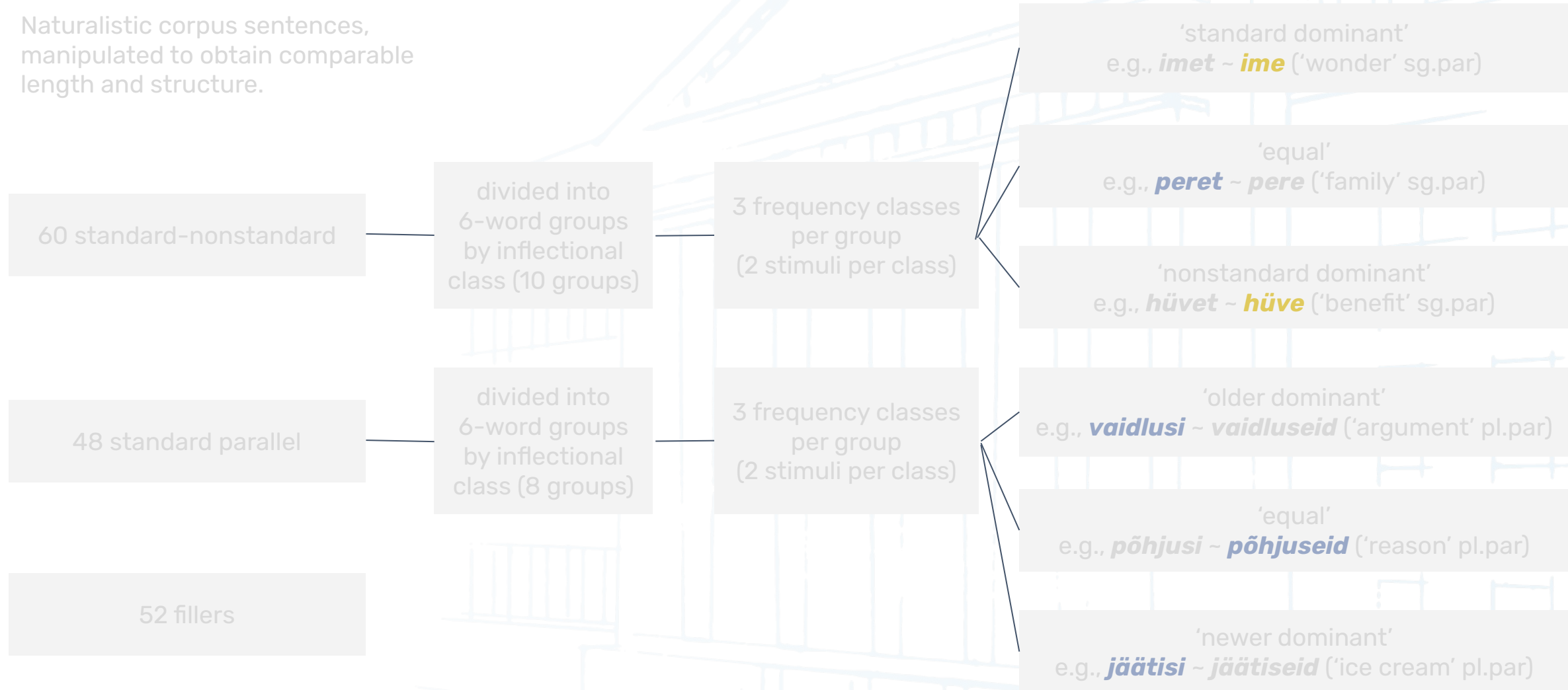
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



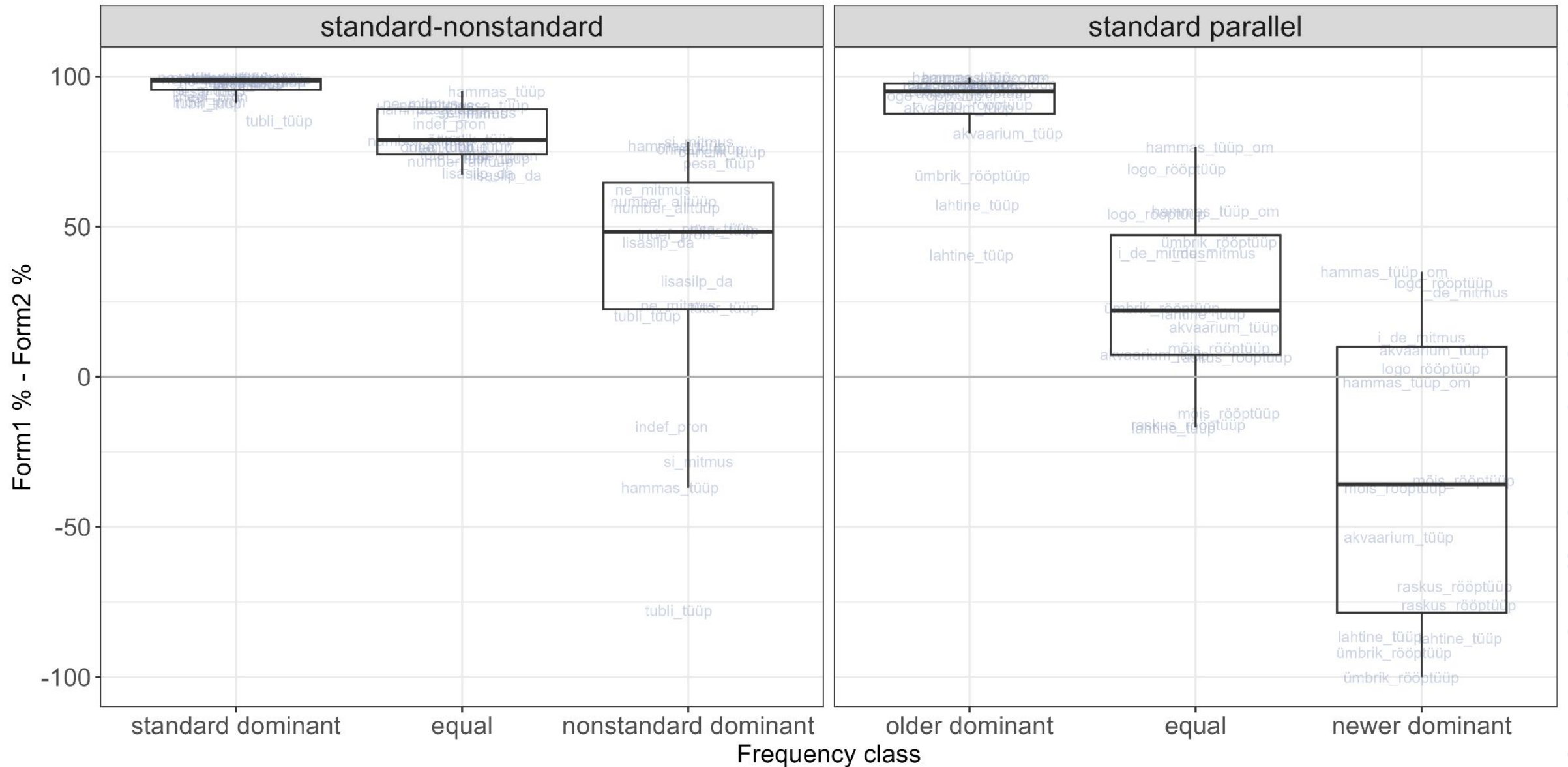
## Study 2: Study design

Based on frequencies from  
the Estonian National Corpus (2023),  
3.8 billion tokens

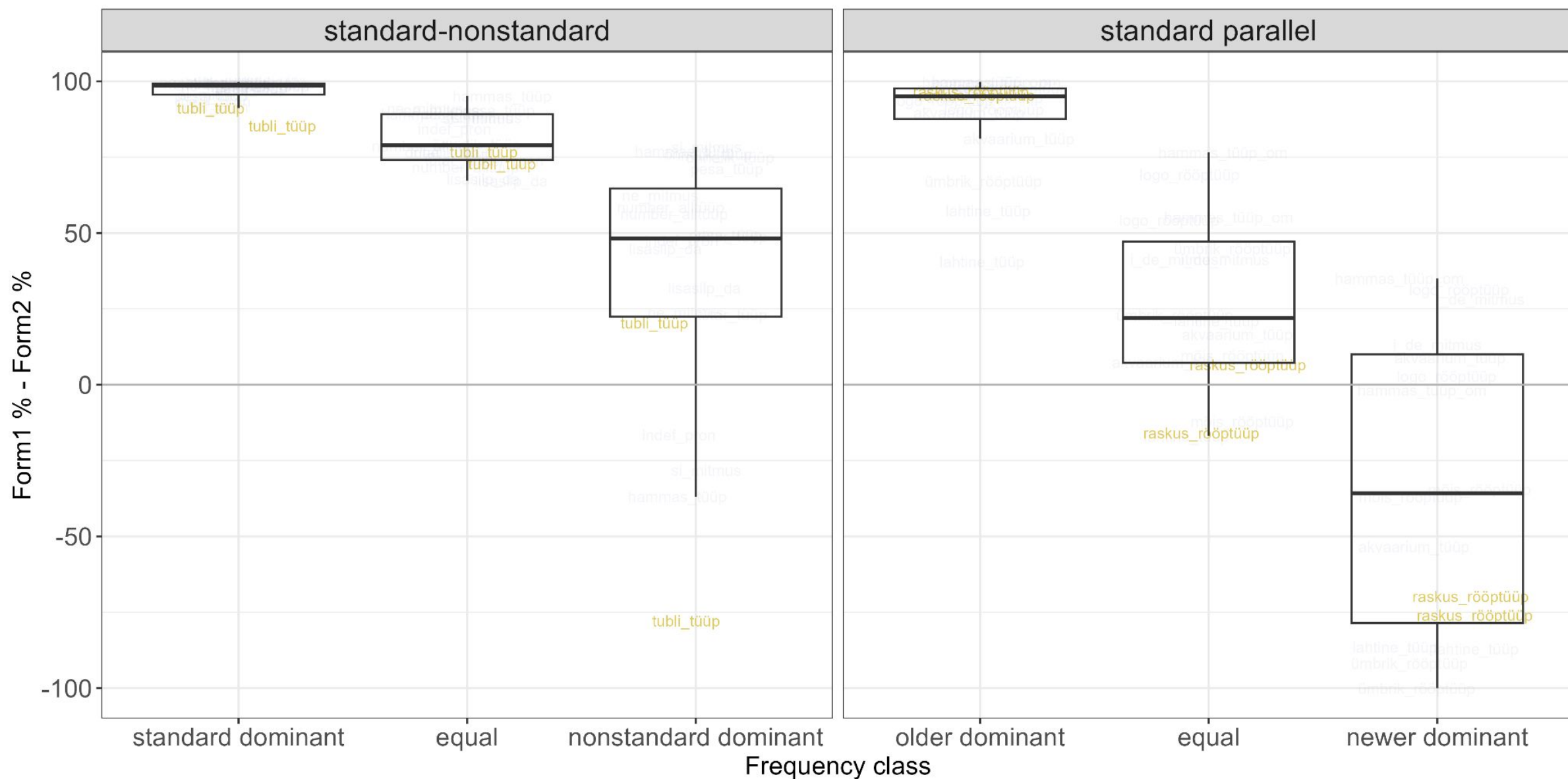
Naturalistic corpus sentences,  
manipulated to obtain comparable  
length and structure.



## Study 2: Distribution of the stimuli

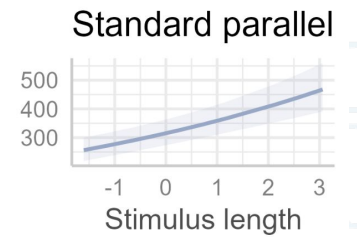
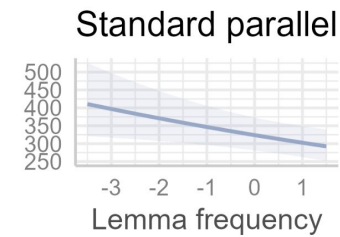
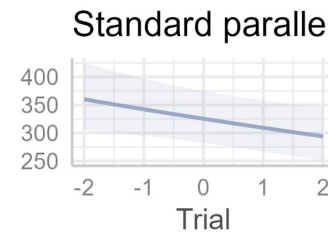
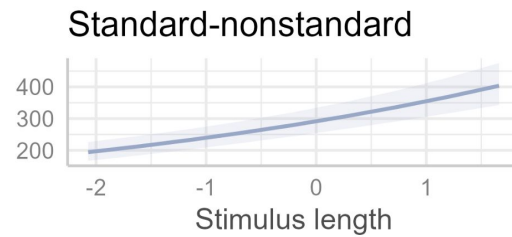
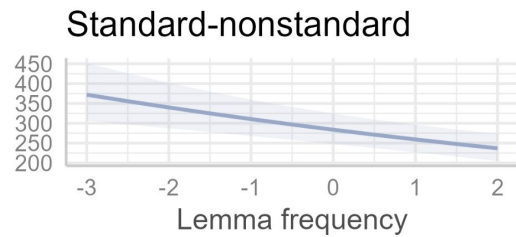
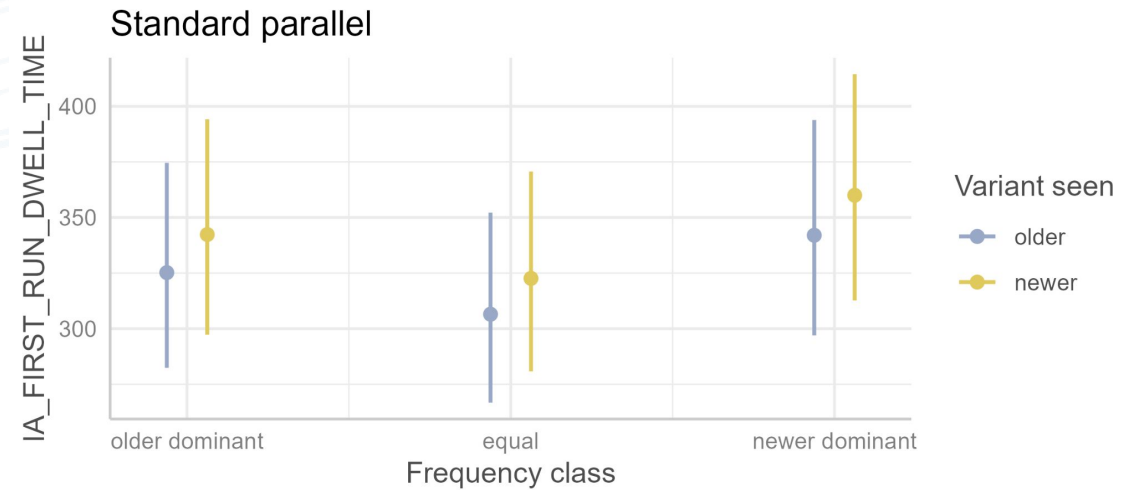
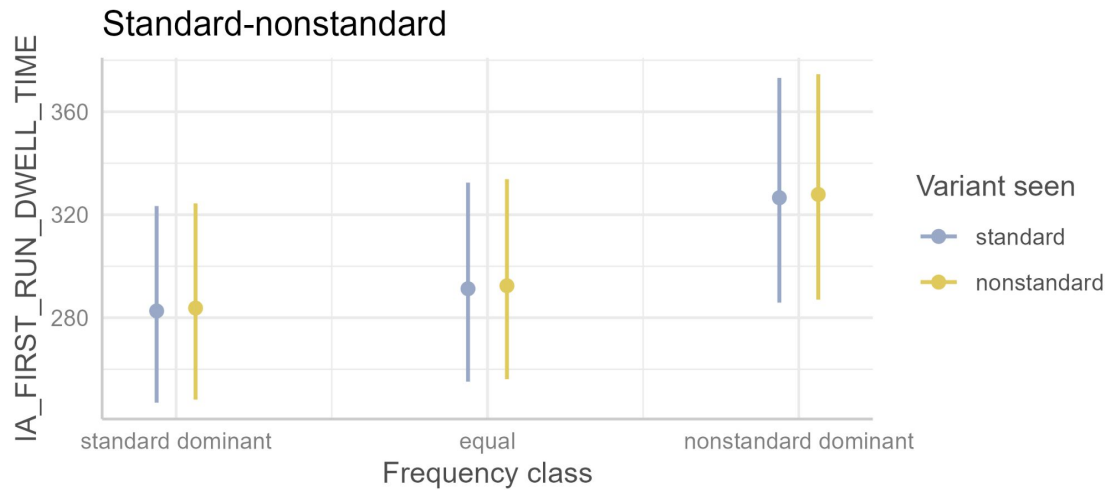


## Study 2: Distribution of the stimuli

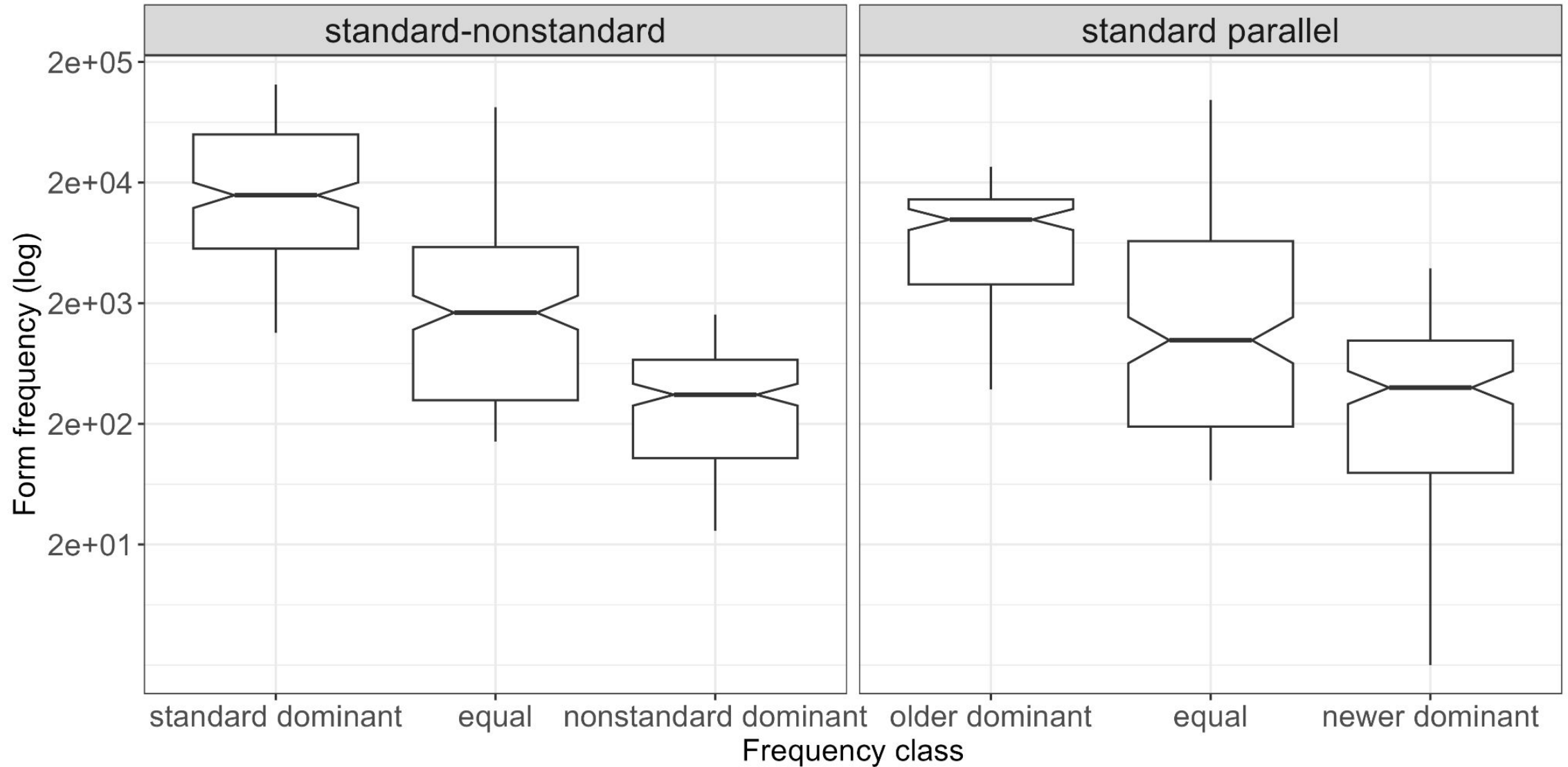


# Study 2: First pass reading (gaze duration)

First run dwell time ~ Variant seen + Frequency class + Trial index + Position in the sentence (log) + Stimulus length + Lemma frequency (log)

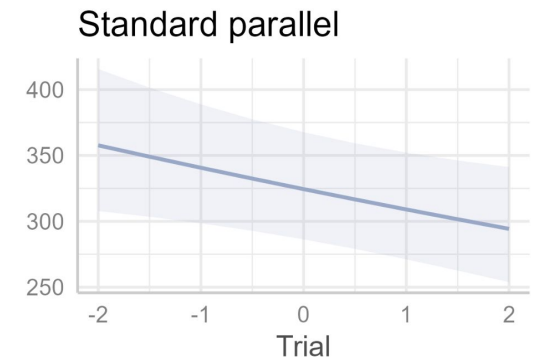
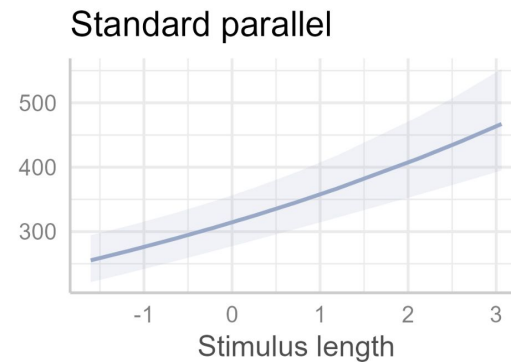
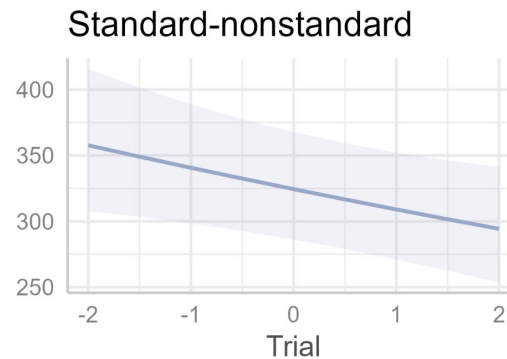
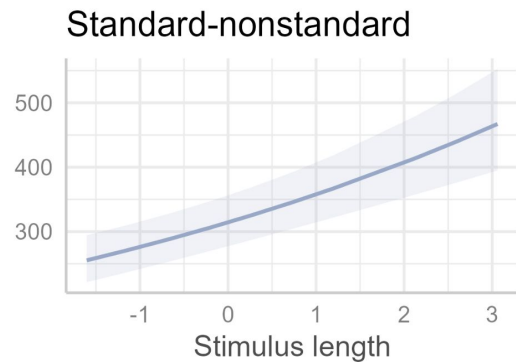
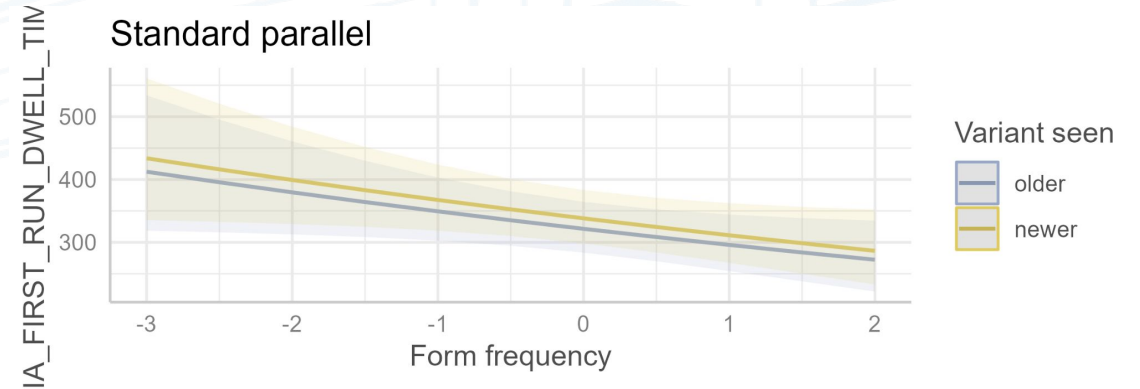
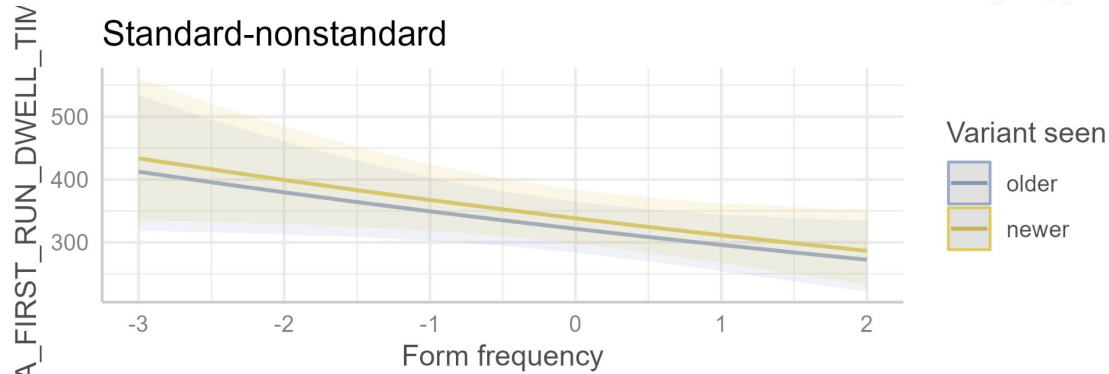


## Study 2: First pass reading (gaze duration)



# Study 2: First pass reading (gaze duration)

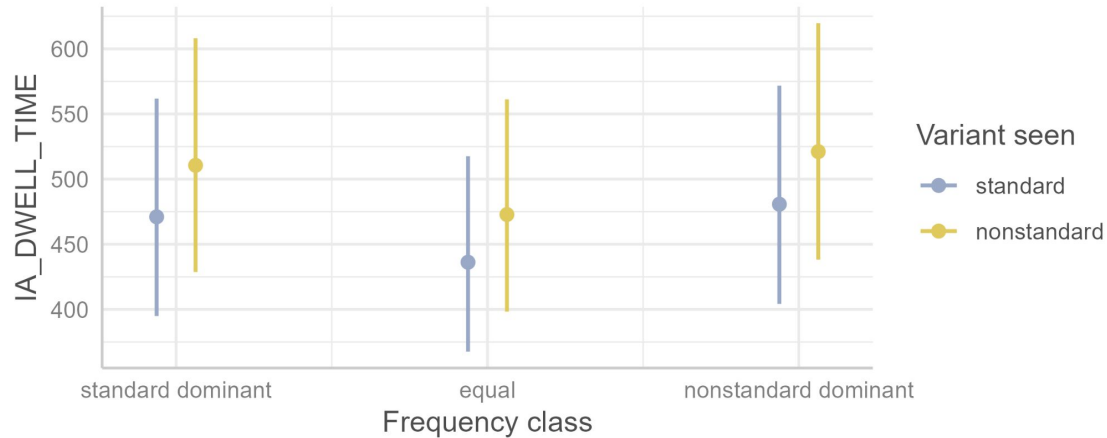
First run dwell time ~ Variant seen (**- Frequency class**) + **Form frequency (log)** + Trial index + Position in the sentence (log) + Stimulus length + Lemma frequency (log)



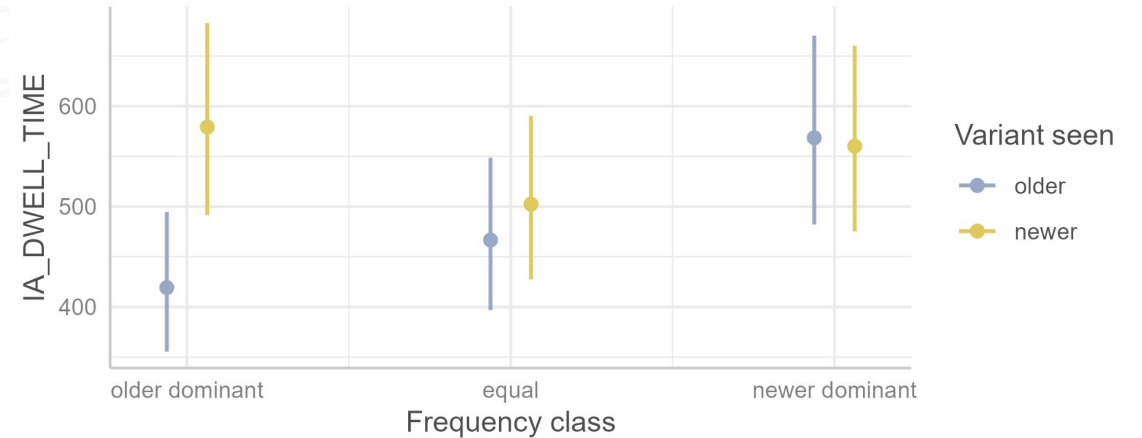
# Study 2: Total dwell time

Total dwell time ~ Variant seen + Frequency class + (Variant seen : Frequency class) + Trial index + Position in the sentence (log) + Stimulus length + Lemma frequency (log)

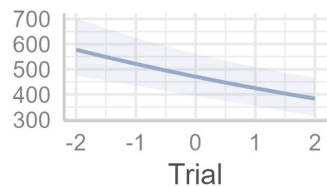
Standard-nonstandard



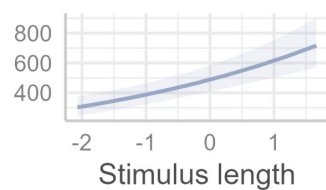
Standard parallel



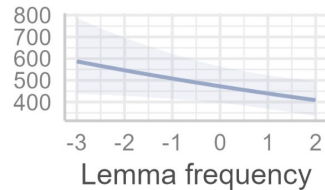
Standard-nonstandard



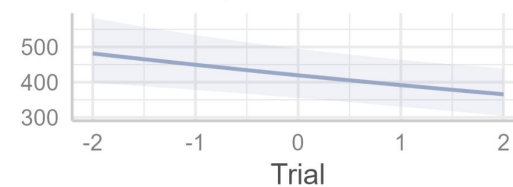
Standard-nonstandard



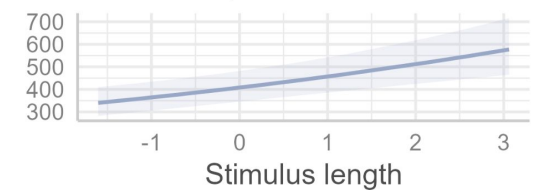
Standard-nonstandard



Standard parallel



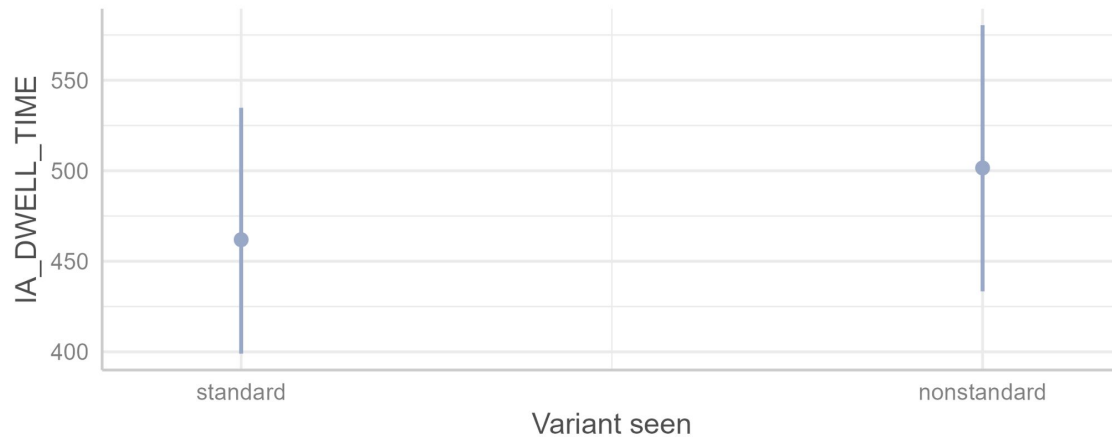
Standard parallel



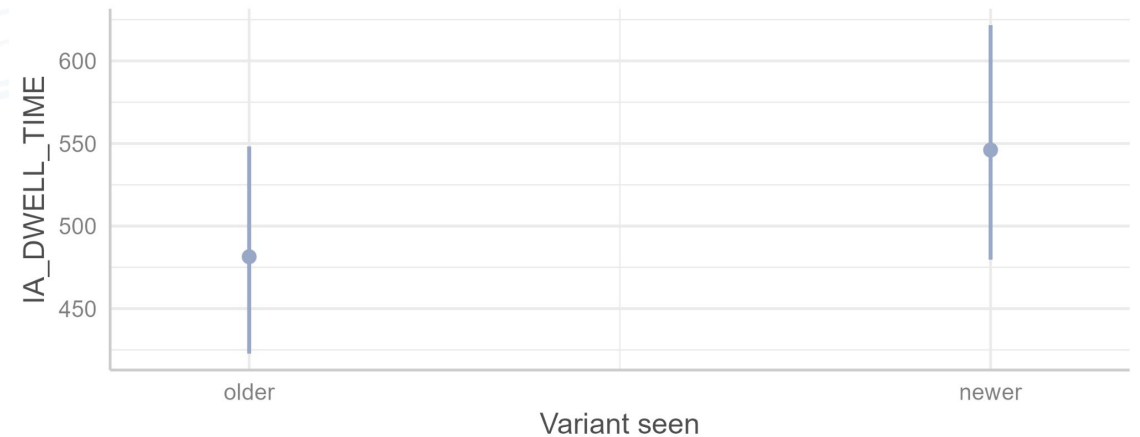
## Study 2: Total dwell time

Total dwell time ~ Variant seen (**- Frequency class**) + **Form frequency (log)** + Trial index + Position in the sentence (log) + Stimulus length + Lemma frequency (log)

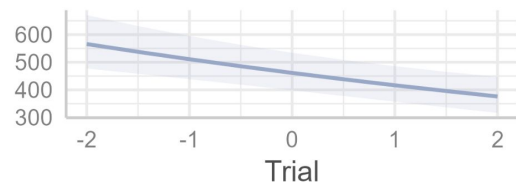
Standard-nonstandard



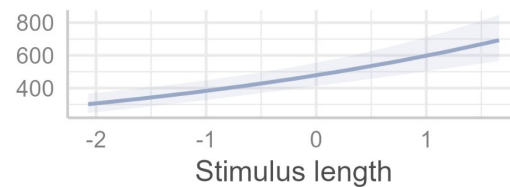
Standard parallel



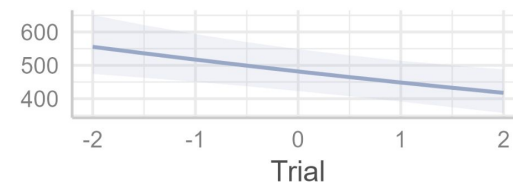
Standard-nonstandard



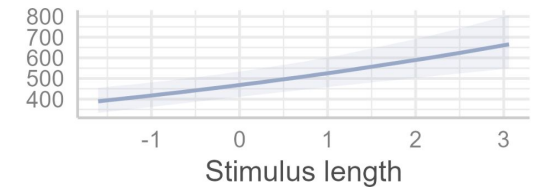
Standard-nonstandard



Standard parallel



Standard parallel



# Summary of study 2

Based on the small pilot study with 10 (language-aware!) participants,

- **variation does not appear to be a problem** for lexical and contextual comprehension in reading. Rather, it is frequency which affects the first pass reading;
- participants seem to be **sensitive to the non-conventionalised status** of word forms (such forms are more often reread). However, this appears to apply to standard and nonstandard forms alike.

# Thank you!

