

# Automatic Extraction of Language-Specific Biomarkers of Healthy Aging In Icelandic

Elena Callegari, Iris Nowenstein, Ingunn Jóhanna Kristjánsdóttir & Anton Karl Ingason

# Language & Alzheimer's Disease

- **Alzheimer's Disease affects what we say and how we say it (Ahmed et al. 2013; Croisile et al. 1996; Kavé and Dassa 2018, i.a.).**
- **Automatic language analysis offers a promising diagnostic and monitoring tool:**
  - **Non-invasive:** far less intrusive than traditional methods (e.g. lumbar punctures).
  - **Cost-effective:** more affordable than e.g. MRI.
  - **Convenient:** could be integrated into remote assessments (e.g., smartphone apps).

# Language & Aging

- **Need for baseline:** what constitutes 'normal' speech patterns in older, healthy individuals?
- **Speech patterns naturally change with age** (Bortfeld et al. 2001; Bóna 2014; Luo et al. 2019; Cho et al. 2021; Spieler and Griffin 2006; Jacewicz et al. 2010; Kemper et al. 2003, i.a.)
- **Differentiating disease-related changes from age-related changes is crucial.**



UNIVERSITY  
OF ICELAND

# Questions

# Objectives

- **To establish a linguistic baseline for what represents healthy aging in older Icelandic speakers using automatically extracted PoS rates.**
- **To understand the influence of different speech elicitation tasks on PoS rates.**

# Methods

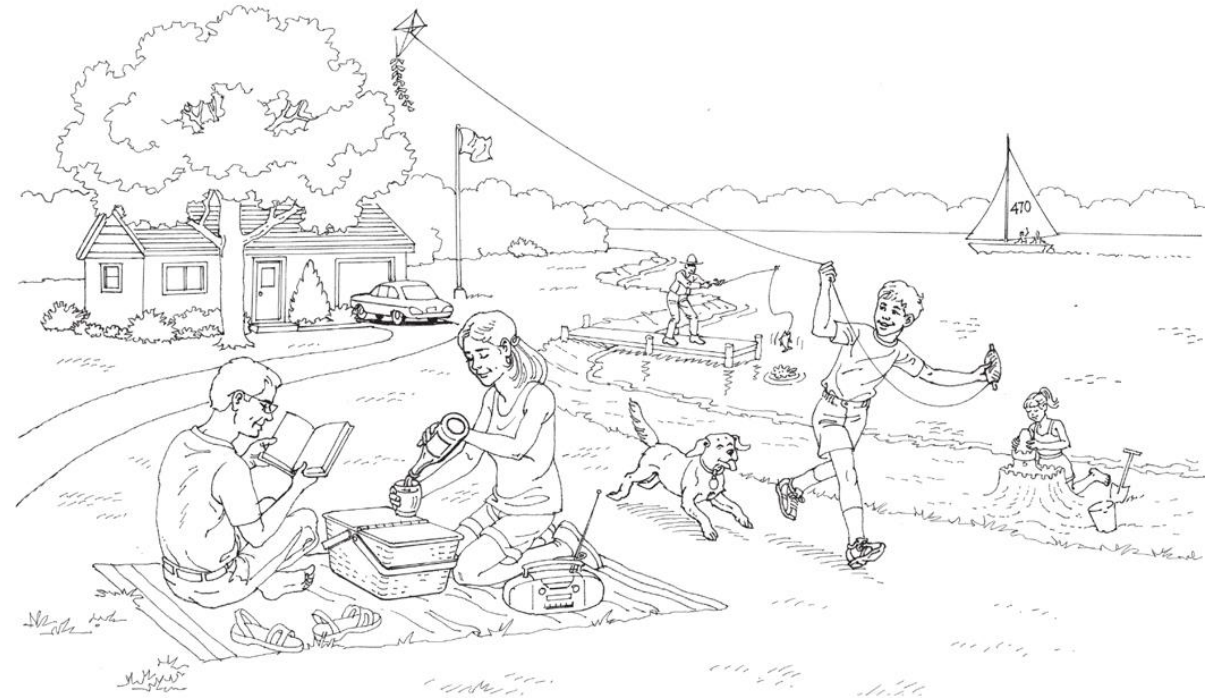
# Participants

- **30 participants (15 females).**
- **Age range: 60 to 80.**
- **Participants were asked to answer 3 questions while being audio-recorded.**
- **Speech samples were transcribed manually by trained annotators.**

# Speech Protocol

Each participant was asked to describe:

- The “picnic scene” (*Western Aphasia Battery Revised*, Kertesz, 1982).
- How they would plan a trip to Akureyri.
- Their childhood home.





# Transcription Processing Methods

- Text was tokenized using *Tokenizer* (Þorsteinsson et al., 2022), a tool for Icelandic text tokenization.
- PoS tags were generated using *GreynirSeq*, a natural-language-parsing toolkit for Icelandic.

## Features Extracted

1. Rate of nouns
2. Rate of pronouns
3. Rate of adverbs
4. Rate of conjunctions
5. Rate of verbs
6. Rate of inflected verbs
7. Rate of past participles
8. Rate of subjunctives
9. Rate of prepositions
10. Rate of DPs with dative case
11. Rate of DPs with genitive case
12. Type-token ratio
13. Rate of unfinished words
14. Rate of corrections

- We used linear mixed effects models to analyze normalized language features.
- Language features were normalized based on relevant counts:
  - General features: total number of intelligible words spoken.
  - Specific grammatical categories (like past participles or subjunctive): number of verbs.
  - Case marking features: number of words that had case marking.
- Variables:
  - Fixed effects: task type, participant age, and total word count.
  - Random effects: individual differences among participants.
- Models built progressively, adding task type, age, and task\*age interaction.
- Best model selected using Likelihood Ratio Test.

# Results

# Effect of Task Type

- **Significant improvements to the fit of the model when adding task type for 11 out of 14 variables tested.**
- **Variables that do not show task-type effects:**
  - **Rate of pronouns**
  - **Rate of adverbs**
  - **Rate of prepositions**

# Effect of Age

- **Age effect much less robust than task-type effect.**
- **3 out of 14 variables only show an age effect:**
  - **Rate of adverbs**
  - **Rate of verbs**
  - **Dative case for age\*task interaction**

# Discussion

# Pronouns, Nouns and Pronoun-Noun Ratios

## Our study:

- Rate of pronouns **NOT** task-sensitive.
- Pronoun-noun ratio **IS** task-sensitive.
- Highest noun rate: picture-description task, lowest: trip-planning.

## Existing research:

- Increased use of pronouns in English speakers with Alzheimer's Disease (Petti et al. 2020, Robin et al. 2021, Cho et al. 2022).
- Older (52 to 89) speakers of English use more pronouns than younger (18 to 22) speakers (Cho et al. 2021)
- Results usually determined on the basis of picture description tasks, with pronoun-to-noun ratios sometimes being computed (Petti et al. 2020).



# Age Effects

- **Age effect much less robust than task-type effect.**
- **Expected result considering lack of contrast to younger speakers.**
- **Interaction with task type for dative case marking is one of the variable affected by age; this can only be studied in morphologically richer languages like Icelandic.**

# Summary and future directions

- We administered three language tasks to participants aged 60–80: picture description, trip planning, and description of one's childhood home.
- Task effect for 11 out of 14 linguistic variables studied, highlighting the substantial influence of sampling methods on language production.
- Among the variables showing statistically significant task effects, rate of the genitive and rate of subjunctive verbs.
- Aging effects more subtle, being evident in 3 of the 14 variables, including an interaction with task type for dative case marking.
- Our findings underscore the need to examine languages other than English to fully understand the effects of aging on language production.

**Takk fyrir að hlusta!**

**Thank you for your attention!**