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

Elena Callegari, Iris Edda Nowenstein, Ingunn J. Kristjánsdóttir, Anton Karl Ingason
University of Iceland,
Reykjavík, Iceland
{ecallegari, irisen, ijk4, antoni}@hi.is

Abstract

This study examines the influence of task type and healthy aging on various automatically extracted part-of-speech features in Icelandic. We administered three language tasks to participants aged 60–80: picture description, trip planning, and description of one's childhood home. Our findings reveal significant task effects on 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, we find the rate of the genitive and subjunctive, variables which can only be studied in morphologically richer languages like Icelandic. On the other hand, rates of pronouns, adverbs, and prepositions remained stable across task types. Aging effects were more subtle, being evident in 3 of the 14 variables, including an interaction with task type for dative case marking. These findings underscore the significance of task selection in studies targeting linguistic features but also emphasize the need to examine languages other than English to fully understand the effects of aging on language production. Additionally, the results have clinical implications: understanding healthy aging's impact on language can help us better identify and study changes caused by Alzheimer's Disease in older adults' speech.

Keywords: Alzheimer's Disease, Icelandic, Part-of-Speech Rates, Healthy Aging, Language Biomarkers

1. Introduction

Alzheimer's Disease (AD) affects what we say and how we say it. Multiple studies have shown that individuals suffering from AD exhibit difficulties with word retrieval (Croisile et al., 1996; Kavé and Dassa, 2018), produce fewer information units and content words (Ahmed et al., 2013; Croisile et al., 1996; Kavé and Dassa, 2018), and use more pronouns than healthy age-matched controls (Kavé and Dassa, 2018).[1] Language changes are already detectable when individuals are diagnosed with Mild Cognitive Impairment (Kavé and Dassa, 2018), a stage of the disease that can occur up to 8 years before the onset of mild Alzheimer's disease, and potentially even before that (Ahmed et al., 2013; Forbes-McKay and Venneri, 2005; Garrard et al., 2005). Spoken language can thus be of use in the measuring of neurological health and diagnosis of early-stage AD.

Recent advancements in Natural Language Processing (NLP) have sparked interest in the possibility of using automatic language analysis as an affordable, non-invasive, quick method which can contribute to the diagnosis of AD as well as monitor its progression (Clarke et al., 2020; de la Fuente Garcia et al., 2020; Callegari et al., 2023). The main procedures currently available to diagnose AD include cognitive tests in combination with PET or MRI, and/or the sampling of cerebrospinal fluid through lumbar punctures. However, these procedures are costly and often have long waiting times. Automatic language analysis is both less intrusive and considerably less costly than these existing methods, and could be integrated into remote assessment, such as via a smartphone app, further increasing its diagnostic power.

However, to harness the full potential of NLP tools for AD diagnosis, it is imperative to be able to differentiate between Alzheimer's-related speech changes and those naturally occurring with age. Several studies have shown that speech patterns change with age (Bortfeld et al., 2001; Bóna, 2014; Moscoso del Prado Martín, 2017; Luo et al., 2019; Cho et al., 2021; Spieler and Griffin, 2006; Martins and Andrade, 2011; Jacewicz et al., 2010; Kemper et al., 2003): factors such as the relative and absolute frequency of different Part-of-Speech (POS) categories, our speech rate and the number of pauses we produce per minute naturally change as we grow older. Developing effective automatic language analysis tools for AD thus hinges on possessing a well-defined understanding of what constitutes 'normal' speech patterns in age-matched healthy controls: without a solid baseline, it becomes more challenging to determine which changes are typical of aging, and which might indicate the onset or presence of Alzheimer's, potentially increasing the number of

[1]This research is funded by grant nr. 200121-5303 from the Strategic Research Program in Language Technology of the Icelandic Centre for Research.

false positives.
In light of these considerations, in this study we examined the speech production of 30 healthy Icelandic individuals aged between 60 to 80. Our first objective was to establish a baseline of what qualifies as 'normal' or typical speech characteristics for this specific age bracket across a number of language variables, i.e. to establish a *healthy aging language baseline* for Icelandic. Setting this baseline holds significance not just for the development of NLP tools aimed at monitoring and diagnosing AD, but can also be instrumental for Icelandic physicians, neuropsychologists and speech-language pathologists who are assessing the speech and language production of older adults for diverse medical conditions. Presently, Icelandic healthcare professionals lack a comprehensive understanding of what speech and language changes come with age, which limits their ability to objectively evaluate language production in senior Icelandic populations.
Our second objective was to determine how different speech-elicitation tasks affect extracted linguistic features, e.g., whether factors such as the rate of adverbs or pronouns are significantly affected by the type of task that is used to elicit a language sample. This is a second, crucial component for the development of effective automatic language analysis tools for clinical purposes: by understanding the impact of various speech-elicitation tasks on linguistic variables, we can strategically select the task that best accentuates features of clinical relevance for AD when collecting data from users.
This research is particularly novel in that this is the first study of this sort that focuses specifically on Icelandic – a Germanic language spoken by fewer than 400,000 individuals. Our study not only contributes to the understanding of healthy aging language patterns within this particular language group, but also offers the opportunity to explore the cross-linguistic effects of aging on speech. By considering a larger sample of languages, we can determine whether the effects of aging on different linguistic variables are consistent across different languages, or whether there are nuances and variations that are distinctly language-specific. This study, therefore, plays a role in both broadening our comprehension of age-related linguistic changes and in highlighting the importance of considering language-specific variables when developing NLP tools for diverse linguistic communities.

2. Data Collection

2.1. Participants

For this study, we recruited 30 individuals, evenly split between 15 males and 15 females. All participants were between the ages of 60 and 80. The exclusion criteria were: a primary diagnosis of depression of moderate or severe degree, bipolar disorder, schizophrenia, a previous physical brain injury, a neurological disorder or other serious medical condition, a personal history of drug addiction within the past 20 years, issues with alcohol addiction within the past 20 years, the use of antidepressants and the use of benzodiazepine-based medications.
The participants' education levels varied: four held PhDs, nine had Bachelor's degrees, five possessed Master's degrees, five completed mandatory elementary education, three underwent apprenticeships, and two had passed the Icelandic matriculation examination. One participant's educational background was not provided.

2.2. Protocol

Each participant was asked to describe in detail: (i) the "picnic scene" from the Western Aphasia Battery Revised (Kertesz, 1982). This is a black-and-white depiction of a picnic by the lake; (ii) how they would plan a trip to Akureyri, a town in the north of Iceland; (iii) their childhood home. We decided to include more than the traditional picture-description task, used in many studies on AD, because of evidence that picture-description tasks may not accurately reflect the conversational abilities of individuals (Sajjadi et al., 2012a).
The order in which the three main prompts were presented was rotated across participants to mitigate the effect of fatigue on verbal performance. During interviews, participants were encouraged to speak freely and uninterrupted while being audio-recorded.
Speech samples were transcribed manually by trained annotators using transcription methods and guidelines from *Linguistic Data Consortium* at the University of Pennsylvania (Glenn et al., 2010) (see also Callegari et al. (2023) for a detailed overview of how the manual transcriptions were carried out). The transcriptions contain speech from both speakers, i.e. interviewer and interviewee, and accurately annotate any interjections or overlaps, providing detailed transcriptions of the conversations as a whole.
The transcriptions are verbatim and orthographic using standard Icelandic spelling. Filled pauses, false starts, repeated words, repairs, restarts, partial words, spoonerisms, speech errors and speaker noises were all marked and annotated in accordance with the transcription protocol. We followed the LDC guidelines as much as possible with some modifications for Icelandic. These adjustments primarily involve Icelandic discourse particles, which differ from those in English. For instance, we created a list of Icelandic-specific discourse particles, including "uu", "ömm", "sko", and

Variable
rate of nouns
rate of pronouns
rate of adverbs
rate of conjunctions
rate of verbs
rate of inflected verbs
rate of past participles
rate of subjunctives rate
rate of prepositions
rate of DPs with dative case
rate of DPs with genitive case
type-token ratio
rate of unfinished words
rate of corrections

Table 1: Examined Features

"hérna".
Our study received approval from the Icelandic Research Ethics Committee (*Vísindasiðanefnd*) in September 2021.

3. Data Analysis

We took the transcriptions generated from the speech samples collected for each of our 30 participants and processed them to extract specific linguistic features from the participants' language production. The selection of extracted features was based both on previous literature on automatic analysis of linguistic markers of both aging and neurodegeneration (Petti et al. (2020), Robin et al. (2021), Cho et al. (2021), Cho et al. (2022)) as well as properties of Icelandic which are understudied in the field of clinical linguistic markers, where the predominance of English is well-established (e.g. García et al. (2023)). The variables we computed are listed in Table 1.

3.1. Feature Extraction

To extract part-of-speech (POS) rates from the transcriptions, we used the POS tagging functionality of GreynirSeq, a natural-language-parsing toolkit for Icelandic focused on sequence modeling with neural networks (Símonarson et al., 2022). The POS tagger was trained on the Tagged Icelandic Corpus (MIM-GOLD) dataset (Barkarson et al., 2021) on top of IceBERT, an Icelandic BERT-based language model, achieving 98.2% accuracy. We developed a Python program that utilized Tokenizer (Þorsteinsson et al., 2022), a tool for Icelandic text tokenization, to automatically tokenize text in the transcripts. The same program also employed GreynirSeq to annotate the part-of-speech (POS) tag for each token. The number of hits in each POS category was counted for every participant and task, as well as counts for inflected verbs, past participles, verbs in the subjunctive, words in the nominative, accusative, dative and genitive, total word count and type-token ratio (moving average). The type-token ratio was calculated using a window size of 100, following the method proposed by Covington and McFall (2010). While we also counted instances of unfinished words and corrections, note that we specifically excluded these from the counts of the other POS categories, though we did note their frequencies separately.

3.2. Statistical Models

To analyze the results, we ran linear mixed effects models (Bates et al., 2015) with our normalized language features as the outcome variable. Features were either normalized based on the total number of intelligible words or the total number of words in specific POS categories, with the past participles, inflected verbs and subjunctive being normalized based on the number of verbs, while the case marking features were normalized based on the number of case marked words. The sample type, participant age and total word count were the explanatory variables (fixed effects) of the models and we included random intercepts and slopes by participant. We conducted a nested model comparison (Likelihood Ratio Test) by progressively adding to a base model with random effects in the following order: 1) Task Type, 2) Age and 3) Task type * Age Interaction. This constituted an analysis with four models which were compared for each variable. The dataset and R script have been made available through the Open Science Framework.[2]

4. Effect of Task Type

We first show how the different POS rates, the type-token ratio (moving average), the rate of unfinished words and corrections are affected by task type. Recall that we had three types of language-sampling tasks: i) picture description, ii) planning of a trip, and iii) description of one's childhood home.
Asking participants to describe a picture scene is a commonly used method to elicit speech samples. A particularly common picture in this respect is the "Cookie Theft" picture (Goodglass et al., 1983). An alternative approach consists in asking participants to recount a narrative that is presented in pictures, such as the "Frog, Where Are You?" (Mayer, 2003) story. Picture-description tasks have been widely adopted because of their simplicity and standardization. Moreover, there exist large available datasets (such as MacWhinney (2019)'s TalkBank) that were created using

[2] https://osf.io/53mhv/?view_only=585d6c8cb95b478e84e8eaceb5629ca9

Variable	Task type	Chisq.
nouns	yes: p < 0.001	40.56
pronouns	no: p = 0.26	2.65
adverbs	no: p = 0.29	2.56
conjunctions	yes: p < 0.01	9.31
verbs	yes: p < 0.001	25.57
inflected verbs	yes: p < 0.001	59.92
participle rate	yes: p < 0.01	11.61
subjunctive rate	yes: p < 0.001	35.01
prepositions	no: p = 0.45	1.98
dative	int.: p < 0.01	10.7
genitive rate	yes: p < 0.001	23.96
type-token ratio	yes: p < 0.05	6.73
unfinished words	yes: p < 0.05	7.07
corrections	yes: p < 0.001	14.48

Table 2: Model fit improvement when adding the task type variable (Likelihood Ratio Test).

picture descriptions as the chosen method, allowing one to compare one's results with existing ones collected for other participants and languages. At the same time, picture descriptions pose a series of drawbacks: the elicited speech sample is often quite short, and features limited lexical and syntactic richness (Ash et al., 2006). Moreover, the task itself is not particularly representative of everyday speech.

Limited studies exist that compare the effectiveness of different speech-sampling methods in detecting the early stages of Alzheimer's Disease (AD). Findings suggest that conversation via semi-structured interviews and picture descriptions generate different error types (Sajjadi et al., 2012b), and that task nature influences machine learning classification accuracy in distinguishing between patients and controls (Beltrami et al., 2016; Clarke et al., 2021). For example, Clarke et al. (2021) explored linguistic feature-based classifications of discourse from 50 participants (25 healthy, 25 with mild Alzheimer's Disease (AD) or Mild Cognitive Impairment (MCI)) across five different speech tasks.

The authors show that the choice of speech task impacts the performance of classifiers trained to recognize mild AD and MCI: classifiers reach an overall accuracy of 78% when participants are asked to narrate the Cinderella story, but only 62% when participants were asked to narrate the "Frog, Where Are You?" (Mayer, 2003) novel, a story with which they were unfamiliar.

Our results for task effect are illustrated in Table 2 and Figures 1, 2 and 3. We found significant improvements to the fit of the model when adding task type for 11 out of 14 variables tested, which represents additional evidence in support of the idea that the type of task used can significantly affect the linguistic composition of the analyzed sample. Interestingly, the only variables that do not show task type effects are the rate of pronouns, the rate of adverbs, and the rate of prepositions.

This is clinically relevant: as an example, there is ample evidence for increased use of pronouns in English speakers with Alzheimer's Disease (e.g. Petti et al. (2020), Robin et al. (2021), Cho et al. (2022)) as well as work (Cho et al., 2021) showing that older ($52-89$ old) speakers of English use more pronouns as compared to younger ($18-22$ years old) speakers of English. This work is usually conducted based on picture descriptions (e.g. Robin et al. (2021), Cho et al. (2021), Cho et al. (2022)), with pronoun-to-noun ratios sometimes being computed (Petti et al., 2020).

Our results suggest that the pronoun rate as a measure can be robust across different language sampling tasks for Icelandic, but that the pronoun-noun ratio is task-sensitive, with individual noun rates fluctuating significantly across tasks. The highest rate of nouns is observed in the picture-description task. This is intuitively aligned with the nature of the task: describing a visual scene naturally requires the use of nouns to identify and discuss various elements within the image. For instance, if the picture features people, objects, and a setting, participants would inherently name these elements, leading to increased noun usage. On the other hand, the trip-planning narrative showed the lowest rate of nouns. This makes sense as planning a trip revolves more around actions, intentions, and sequences of events rather than specific entities. In such narratives, participants are more likely to use verbs to describe activities they would do on the trip and conjunctions to link different events. The emphasis shifts from naming specific objects, as in the picture-description task, to discussing actions and future intentions (which also leads to a significantly increased usage of the subjunctive mood), leading to a reduced reliance on nouns.

This distribution of the results across language sample tasks illustrated in the figures relates to another important result, which is the extent to which these task effects wildly vary across variables. For example, the three verb rate measures show varying patterns of language task type effect, with the overall verb measure mostly showing a contrast between the description of the childhood home (lower rate) and the two other tasks. On the other hand, when looking at the rate of finite (inflected) verbs, all types of language samples differ from each other. This is interesting considering that the rate of tense-inflected verbs can be reduced in (English) neurodegeneration (Cho et al., 2022), but healthy older speakers of English have also been shown to use more verbs than younger speakers (Cho et al., 2021).

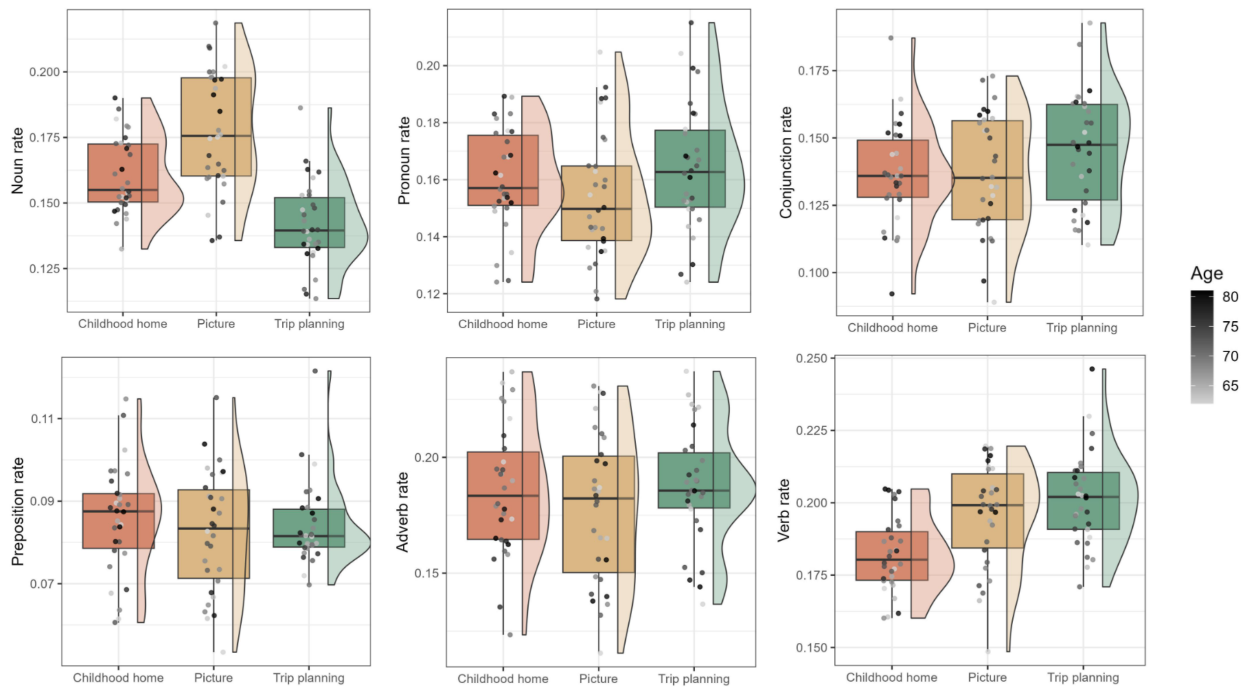

Figure 1: Individual POS rates across types of language sample tasks for features normalized on the total number of words, N = 48. Lighter dots indicate a lower age.

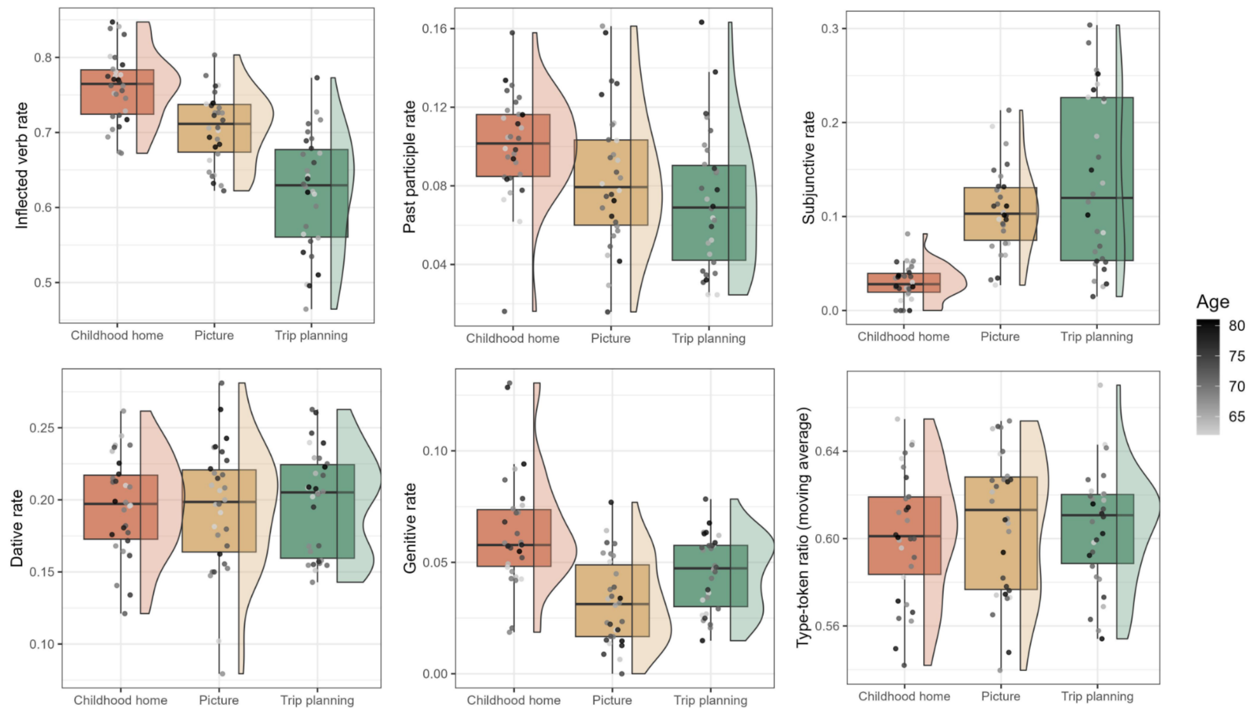

Figure 2: Individual linguistic feature rates across types of language sample tasks for the type-token ratio and features normalized on the total number of verbs (top line) and the total number of case marked words (dative and genitive rate), N = 48. Lighter dots indicate a lower age.

Turning to the variables reflecting characteristics of Icelandic which differ from English, it is inter-

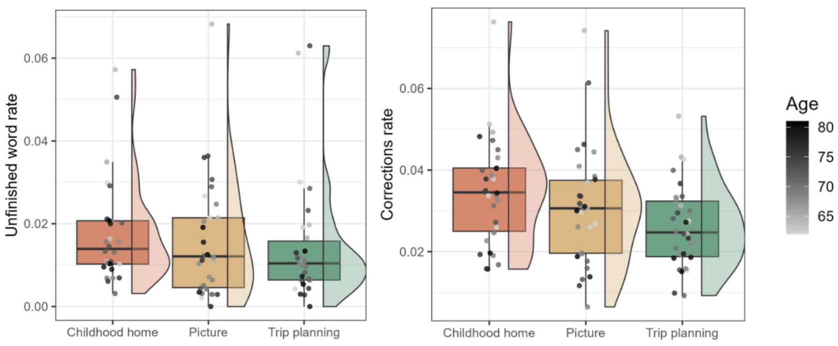

Figure 3: Individual rates of unfinished words and corrections across types of language sample tasks, normalized on the total number of words, N = 48. Lighter dots indicate a lower age.

esting to note the large differences in the rate of the subjunctive across language sample types, with a remarkably low rate of subjunctives used in the description of participants' childhood home, and with high rates when participants discuss the planning of a possible trip. Since the subjunctive has largely disappeared from English (and various other Germanic languages), very little is known about ways in which it could be affected in aging or neurodegenerative disease. Still, work on Greek and Italian speakers with probable Alzheimer's Disease (Fyndanis et al., 2017) suggests that dementia may alter how mood is used in language.

When it comes to case marking, also largely lost in English but preserved in Icelandic (McFadden, 2020), we observe sample type effects as well as the only significant interaction between age and sample type found in the study (for the dative rate). As all the figures have participant age represented as well, it can be seen that older participants use the dative case more for two of the three task types, trip planning and picture description, but not for the childhood home. Although these results are difficult to interpret, it is clear that more detailed work needs to be conducted when it comes to the use of case marking in aging and neurodegeneration. For example, Bose et al. (2021) show that the use of case marking changes in the language of speakers of Bengali who have Alzheimer's Disease.

5. Effect of Age

An important aspect of investigating possible language-specific manifestations of neurodegeneration is to be able to distinguish between changes in healthy aging and changes in speech and language which could be markers of disease. Numerous studies have shown that the language productions of older individuals differ from those of younger individuals (Bortfeld et al., 2001; Bóna, 2014; Moscoso del Prado Martín, 2017; Luo et al., 2019; Cho et al., 2021; Spieler and Griffin, 2006; Martins and Andrade, 2011; Jacewicz et al., 2010; Kemper et al., 2003). For example, Cho et al. (2021) examined the descriptions of the Cookie Theft picture produced by 37 older (age range: 52 to 89) and 76 younger (18 to 22) healthy participants. They found that older speakers produce shorter clauses, more fillers, pronouns and verbs than younger individuals, but use fewer conjunctions, determiners and nouns. They also noticed a correlation between age and vocabulary used, with older speakers exhibiting overall lower lexical diversity than younger participants. However, all the above-cited studies, with the exception of Bóna (2014), based on Hungarian, and Martins and Andrade (2011) on Brazilian Portuguese, were based on English. Bóna (2014) also focused on examining acoustic variables only, such as speech rate, articulation rate, and length of pauses, which are likely variables that are more stable across different languages. Therefore, the effects of healthy aging on morphologically rich languages such as Icelandic, which makes use of a case system, have so far been undocumented.

In addition to comparing older individuals with younger ones, to develop effective clinical NLP applications for AD detection, it is also imperative to look at differences within the older age group. For example, do language patterns vary significantly depending on whether an individual is in their 60s versus their 70s? This is why in our study we specifically investigate age effects in participants between the ages of 60 and 80.

Table 3 illustrates which linguistic variables showed significant model fit improvements when participant age was added. As can be seen, and is additionally illustrated in Figures 1, 2 and 3, aging effects only appear with 3 of the 14 variables tested, showing that the aging effects are much less robust than the effects of task type. This is

Variable	Age	Chisq.
nouns	no: p = 0.68	0.18
pronouns	no: p = 0.48	0.46
adverbs	yes: p < 0.01	6.52
conjunctions	no: p = 0.97	0.01
verbs	yes: p < 0.05	5.45
inflected verbs	no: p = 0.66	0.20
participle rate	no: p = 0.06	3.42
subjunctive rate	no: p = 0.46	0.56
prepositions	no: p = 0.19	1.68
dative	int.: p < 0.01	10.7
genitive rate	no: p = 0.21	1.55
type-token ratio	no: p = 0.06	3.41
unfinished words	no: p = 0.15	2.06
corrections	no: p = 0.06	3.57

Table 3: Model fit improvement when adding the age variable (Likelihood Ratio Test).

to be expected considering the lack of contrast to younger speakers. Nevertheless, the presence of significant effects points to the importance of aging effects within older speakers. The results for the dative have already been described, but it is interesting to see an age effect also emerge in the rate of verbs, with an increased use as participants age. This is comparable to the results of Cho et al. (2021) in their study contrasting younger and older speakers of English. On the other hand, their results did not show an aging effect for adverbs as ours do, but such an effect can still be found when speakers with dementia are compared to healthy controls (Cho et al., 2022).

6. Concluding Remarks

In our research, we analyzed the speech patterns of 30 healthy Icelandic individuals aged between 60 and 80. Our primary objectives were: i) to establish a linguistic baseline for what represents healthy aging in older Icelandic speakers using automatically extracted POS features, and ii) to understand the influence of different speech elicitation tasks on these POS features.

Our exploration into the effects of task type on the different linguistic variables offers insights into the importance of choosing the right language sampling method. Picture description tasks are commonly used in clinical language sampling due to their simplicity, offering both benefits and drawbacks. Although widely adopted, these tasks may produce data that has limited lexical depth. Our findings reveal the pronounced impact of sampling methods on linguistic variables, with 11 of the 14 variables studied showcasing noticeable variation depending on the task type. A particularly interesting finding relates to the rate of pronouns used, which appears to be more or less stable across different task types. This consistency holds clinical significance, especially considering that pronoun rate has been identified as an indicator of Alzheimer's disease in several studies (Kavé and Dassa, 2018; Petti et al., 2020; Robin et al., 2021; Cho et al., 2022). Whereas the pronoun rate remained stable across different elicitation tasks, the pronoun-to-noun ratio did not, as the noun rate was highly dependent on task type, with tasks eliciting descriptions of visual cues resulting in a higher number of nouns being produced across all ages. This suggests that caution should be exercised when computing pronoun-to-noun ratios, an equally popular measure used in clinical linguistic studies focusing on AD. Unless the specific sampling type is taken into account, such computations might lead to skewed results.

Historically, the focus of studies on age effects on language has largely been on mapping contrasts between older and much younger individuals (e.g. individuals in their 30s versus those in their 70s), particularly within the English language domain. Our study ventured into the relatively uncharted territory of aging effects within a narrower age bracket of older individuals, specifically in the context of a morphologically richer language like Icelandic. While we noticed fewer significant results when we looked at age effects, it is interesting to note that even with our concentrated age sample, spanning just 20 years, we identified variables with variation linked to age. This underscores the importance of examining narrower age bands when evaluating language changes in older populations. Moreover, the significant interaction we observed between task type and age for dative usage further shows the need for increased linguistic diversity in research on aging effects in language production.

7. Bibliographical References

Samrah Ahmed, Anne-Marie F Haigh, Celeste A de Jager, and Peter Garrard. 2013. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, 136(12):3727–3737.

Sharon Ash, Peachie Moore, Sameer Antani, Gwyneth McCawley, Melissa Work, and Murray Grossman. 2006. Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology*, 66(9):1405–1413.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Daniela Beltrami, Laura Calzà, Gloria Gagliardi, Enrico Ghidoni, Norina Marcello, Rema Rossini Favretti, and Fabio Tamburini. 2016. Automatic identification of mild cognitive impairment through the analysis of italian spontaneous speech productions. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2086–2093.

Judit Bóna. 2014. Temporal characteristics of speech: The effect of age and speech style. *The Journal of the Acoustical Society of America*, 136(2):EL116–EL121.

Heather Bortfeld, Silvia D Leon, Jonathan E Bloom, Michael F Schober, and Susan E Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, 44(2):123–147.

Arpita Bose, Niladri S. Dash, Samrah Ahmed, Manaswita Dutta, Aparna Dutt, Ranita Nandi, Yesi Cheng, and Tina M. D. Mello. 2021. Connected Speech Characteristics of Bengali Speakers With Alzheimer's Disease: Evidence for Language-Specific Diagnostic Markers. *Frontiers in Aging Neuroscience*, 13.

Elena Callegari, Agnes Sólmundsdóttir, and Anton Karl Ingason. 2023. The ACoDe Project: Creating a dementia corpus for Icelandic. In *Proceedings of CLARIN Annual Conference 2023*, pages 100–105.

Sunghye Cho, Katheryn Alexandra Quilico Cousins, Sanjana Shellikeri, Sharon Ash, David John Irwin, Mark Yoffe Liberman, Murray Grossman, and Naomi Nevler. 2022. Lexical and Acoustic Speech Features Relating to Alzheimer Disease Pathology. *Neurology*, 99(4):e313–e322.

Sunghye Cho, Naomi Nevler, Sanjana Shellikeri, Natalia Parjane, David J Irwin, Neville Ryant, Sharon Ash, Christopher Cieri, Mark Liberman, and Murray Grossman. 2021. Lexical and acoustic characteristics of young and older healthy adults. *Journal of Speech, Language, and Hearing Research*, 64(2):302–314.

Natasha Clarke, Thomas R Barrick, and Peter Garrard. 2021. A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. *Frontiers in Computer Science*, 3:634360.

Natasha Clarke, Peter Foltz, and Peter Garrard. 2020. How to do things with (thousands of) words: Computational approaches to discourse analysis in Alzheimer's disease. *Cortex*, 129:446–463.

Michael A. Covington and Joe D. McFall. 2010. Cutting the gordian knot: The moving-average type–token ratio (mattr). *Journal of Quantitative Linguistics*, 17(2):94–100.

Bernard Croisile, Bernadette Ska, Marie-Josee Brabant, Annick Duchene, Yves Lepage, Gilbert Aimard, and Marc Trillet. 1996. Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain and language*, 53(1):1–19.

Sofia de la Fuente Garcia, Craig W Ritchie, and Saturnino Luz. 2020. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. *Journal of Alzheimer's Disease*, 78(4):1547–1574.

Alexandra Economou, Christopher Routsis, and Sokratis G Papageorgiou. 2016. Episodic memory in Alzheimer's, frontotemporal dementia, and dementia with lewy bodies/parkinson disease dementia. *Alzheimer Disease & Associated Disorders*, 30(1):47–52.

Katrina E Forbes-McKay and Annalena Venneri. 2005. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological sciences*, 26:243–254.

Kathleen C Fraser, Jed A Meltzer, Naida L Graham, Carol Leonard, Graeme Hirst, Sandra E Black, and Elizabeth Rochon. 2014. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *cortex*, 55:43–60.

Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2016. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease*, 49(2):407–422.

Valantis Fyndanis, Dimitra Arfani, Spyridoula Varlokosta, Francesca Burgio, Anna Maculan, Gabriele Miceli, Giorgio Arcara, Fabio Palla, Annachiara Cagnin, Sokratis G Papageorgiou, and et al. 2017. Morphosyntactic production in greek- and italian-speaking individuals with probable alzheimer's disease: Evidence from subject–verb agreement, tense/time reference, and mood. *Aphasiology*, 32(1):61–87.

Adolfo M García, Jessica de Leon, Boon Lead Tee, Damián E Blasi, and Maria Luisa Gorno-Tempini. 2023. Speech and language markers of neurodegeneration: a call for global equity. *Brain*, page awad253.

Peter Garrard, Lisa M Maloney, John R Hodges, and Karalyn Patterson. 2005. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 128(2):250–260.

Mirjam I Geerlings, Cees Jonker, Lex M Bouter, Herman J Adèr, and Ben Schmand. 1999. Association between memory complaints and incident Alzheimer's disease in elderly people with normal baseline cognition. *American Journal of Psychiatry*, 156(4):531–537.

Meghan Lammie Glenn, Stephanie M. Strassel, Haejoong Lee, Kazuaki Maeda, Ramez Zakhary, and Xuansong Li. 2010. Transcription methods for consistency, volume and efficiency. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Harold Goodglass, Edith Kaplan, and Sandra Weintraub. 1983. *Boston naming test*. Lea & Febiger Philadelphia, PA.

Joyce L Harris, Swathi Kiran, Thomas P Marquardt, and Valarie B Fleming. 2008. Communication wellness check-up©: Age-related changes in communicative abilities. *Aphasiology*, 22(7-8):813–825.

Símonarson Haukur Barri, Vésteinn Snæbjarnarson, Pétur Orri Ragnarsson, Haukur Páll Jónsson, Svanhvít Lilja Ingólfsdóttir, and Vilhjálmur Þorsteinsson. 2022. Greynirseq-a natural language processing toolkit for icelandic (v0. 2.0).

Ewa Jacewicz, Robert Allen Fox, and Lai Wei. 2010. Between-speaker and within-speaker variation in speech tempo of american english. *The Journal of the Acoustical Society of America*, 128(2):839–850.

Johanna Jakobsdottir, Sven J van der Lee, Joshua C Bis, Vincent Chouraki, David Li-Kroeger, Shinya Yamamoto, Megan L Grove, Adam Naj, Maria Vronskaya, Jose L Salazar, et al. 2016. Rare functional variant in tm2d3 is associated with late-onset Alzheimer's disease. *PLoS genetics*, 12(10):e1006327.

Frank Jessen, Rebecca E Amariglio, Martin Van Boxtel, Monique Breteler, Mathieu Ceccaldi, Gaël Chételat, Bruno Dubois, Carole Dufouil, Kathryn A Ellis, Wiesje M Van Der Flier, et al. 2014. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & dementia*, 10(6):844–852.

Frank Jessen, Birgitt Wiese, Cadja Bachmann, Sandra Eifflaender-Gorfer, Franziska Haller, Heike Kölsch, Tobias Luck, Edelgard Mösch, Hendrik van den Bussche, Michael Wagner, et al. 2010. Prediction of dementia by subjective memory impairment: effects of severity and temporal association with cognitive impairment. *Archives of general psychiatry*, 67(4):414–422.

Cees Jonker, Mirjam I Geerlings, and Ben Schmand. 2000. Are memory complaints predictive for dementia? a review of clinical and population-based studies. *International journal of geriatric psychiatry*, 15(11):983–991.

Gitit Kavé and Ayelet Dassa. 2018. Severity of Alzheimer's disease and language features in picture descriptions. *Aphasiology*, 32(1):27–40.

Susan Kemper, Ruth E Herman, and Cindy HT Lian. 2003. The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology and aging*, 18(2):181–192.

Bo Seon Kim, Yong Bum Kim, and HyangHee Kim. 2019. Discourse measures to differentiate between mild cognitive impairment and healthy aging. *Frontiers in aging neuroscience*, 11:221.

Minxia Luo, Megan L Robbins, Mike Martin, and Burcu Demiray. 2019. Real-life language use across different interlocutors: A naturalistic observation study of adults varying in age. *Frontiers in psychology*, 10:1412.

Brian MacWhinney. 2019. Understanding spoken language through talkbank. *Behavior research methods*, 51:1919–1927.

Brian MacWhinney, Davida Fromm, Margaret Forbes, and Audrey Holland. 2011. Aphasiabank: Methods for studying discourse. *Aphasiology*, 25(11):1286–1307.

Vanessa de Oliveira Martins and Claudia Regina Furquim de Andrade. 2011. Study of pauses in elderly. *Revista Da Sociedade Brasileira de Fonoaudiologia*, 16:344–349.

Mercer Mayer. 2003. *Frog, where are you?* Penguin.

Thomas McFadden. 2020. *Case in Germanic*, Cambridge Handbooks in Language and Linguistics, page 282–312. Cambridge University Press.

Fermín Moscoso del Prado Martín. 2017. Vocabulary, grammar, sex, and aging. *Cognitive Science*, 41(4):950–975.

Kimberly D Mueller, Bruce Hermann, Jonilda Mecollari, and Lyn S Turkstra. 2018. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of clinical and experimental neuropsychology*, 40(9):917–939.

Anna Nikulásdóttir, Jón Guðnason, Anton Karl Ingason, Hrafn Loftsson, Eirikur Rögnvaldsson, Einar Freyr Sigurðsson, and Steinþór Steingrímsson. 2020. Language technology programme for Icelandic 2019-2023. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3414–3422.

Bart Peintner, William Jarrold, Dimitra Vergyri, Colleen Richey, Maria Luisa Gorno Tempini, and Jennifer Ogar. 2008. Learning diagnostic models using speech and language measures. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4648–4651. IEEE.

Ulla Petti, Simon Baker, and Anna Korhonen. 2020. A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association: JAMIA*, 27(11):1784–1797.

Chengxuan Qiu, Miia Kivipelto, and Eva Von Strauss. 2009. Epidemiology of alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues in clinical neuroscience*, 11(2):111–128.

Jill Rasmussen and Haya Langerman. 2019. Alzheimer's disease–why we need early diagnosis. *Degenerative neurological and neuromuscular disease*, pages 123–130.

Jessica Robin, Mengdan Xu, Liam D. Kaufman, and William Simpson. 2021. Using Digital Speech Assessments to Detect Early Signs of Cognitive Impairment. *Frontiers in Digital Health*, 3.

Seyed Ahmad Sajjadi, Karalyn Patterson, Michal Tomek, and Peter J Nestor. 2012a. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology*, 26(6):847–866.

Seyed Ahmad Sajjadi, Karalyn Patterson, Michal Tomek, and Peter J Nestor. 2012b. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology*, 26(6):847–866.

Daniel H Spieler and Zenzi M Griffin. 2006. The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes*, 21(1-3):291–321.

Höskuldur Thráinsson. 2007. *The syntax of Icelandic*. Cambridge University Press, Cambridge.

Eric Williams, Megan McAuliffe, and Catherine Theys. 2021. Language changes in Alzheimer's disease: A systematic review of verb processing. *Brain and Language*, 223:105041.

8. Language Resource References

Barkarson, Starkaður and Andrésdóttir, Þórdís Dröfn and Hafsteinsdóttir, Hildur and Magnússon, Árni Davíð and Rúnarsson, Kristján and Steingrímsson, Steinþór and Jónsson, Haukur Páll and Loftsson, Hrafn and Sigurðsson, Einar Freyr and Rögnvaldsson, Eiríkur and Helgadóttir, Sigrún. 2021. *MIM-GOLD 21.05*. [link].

Þorsteinsson, Vilhjálmur and Óladóttir, Hulda and Þórðarson, Sveinbjörn and Ragnarsson, Pétur Orri and Jónsson, Haukur Páll and Eyjólfsson, Logi. 2022. *Tokenizer for Icelandic text (3.4.2) (22.10)*. [link].

Símonarson, Haukur Barri and Snæbjarnarson, Vésteinn and Ragnarsson, Pétur Orri and Jónsson, Haukur Páll and Ingólfsdóttir, Svanhvít Lilja and Þorsteinsson, Vilhjálmur. 2022. *GreynirSeq - A Natural Language Processing Toolkit for Icelandic (v0.2.0)*. [link].