Interspeech 2024
1-5 September 2024, Kos, Greece

Modelling Lexical Characteristics of the Healthy Aging Population: A Corpus-Based Study

Han Kunmei[1]

[1]National University of Singapore
kunmei.han@u.nus.edu

Abstract

Language ability at an old age is a balance between preservation and decline. Modelling baseline language variation in normal aging thus is important for our understanding of healthy aging, which can help detect cognitive impairments at the prodromal stage. Large language databases and NLP tools enable us to conduct automated quantitative analysis of natural language data. In this study, we aim to demonstrate that (i) age and sex influence old adults' lexical distribution and lexical concreteness; and (ii) using NLP tools and psycholinguistic metrics to process natural language datasets can help to set a normative benchmark of aging languages.

Index Terms: corpus linguistics, NLP, normal aging, PoS distribution, concreteness

1. Introduction

Research about language changes in aging has been an active area in cognitive aging, often studied with cognitive changes and mechanisms related to neural adjustments and compensation [1]. Aging is not always associated with language decline, since different aspects of linguistic processing vary in their sensitivity to aging [2]. Therefore, modelling baseline language variation in healthy aging speakers is important for understanding cognitive changes in normal aging. By setting a normative benchmark of aging language, researchers can evaluate language performance of old adults and identify potential cognitive impairments from their language use.

This study introduces a new way of evaluating elderly speech by adopting NLP tools and psycholinguistic metrics to natural speech data. Specifically, we conduct a year-by-year analysis of lexical characteristics of adults age above 60 years old with natural speech data, and incorporate sex and age as main factors. In Section 2, we briefly review literatures about language changes in comprehension and production, from the word level to the discourse level. Then we focus on language production at the word level and introduce two lexical measures, PoS distribution and lexical concreteness. In Section 3, we introduce the speech data used in this study and the processing method of different lexical measures. Section 4 reports the results. In Section 5, we discuss our findings regarding existing results. We aim to demonstrate that (i) age and sex influence lexicon uses in terms of lexical distribution and lexical concreteness; and (ii) applying NLP tools and psycholinguistic metrics to natural speech can help to set a baseline of normal aging language.

2. Related Work

2.1. Language Variances in Aging

2.1.1. Language Comprehension

Despite the structural decline in language-related brain regions, the ability of language comprehension remains relatively stable in a normal aging process [2]. At the lexical level, both old and young adults adopt similar systems to support single word processing, although older adults may use a more extensive set of brain regions than younger adults [3]. Vocabulary diversity is found to be maintained or even increases as people age [4]. At the sentence level, comprehension requires the integration of different aspects of linguistic knowledge beyond lexicons, such as syntax, semantics, and pragmatics, to construct a representation of the sentence meaning [5]. Complex syntactic structures often pose greater challenge to older adults [6], but sentential contexts provide information to aid old adults' processing of language [7]. At the discourse level, comprehension demands more working memory, since it requires maintaining thematic information over multiple sentences. Older adults are found to perform better in discourse-level comprehension than young adults due to increased experiences and real-world knowledge, which enables efficient comprehension with heuristics [8].

2.1.2. Language Production

Contrasting with the relatively preserved language comprehension ability, old adults are more likely to encounter difficulty in language production. The increased occurrence of the tip-of-the-tongue state (TOT) in the aging process is one of the strongest pieces of evidence for phonological retrieval deficits in old adults, given that a person can produce semantic and grammatical information about the TOT target but only partial information about the phonology of the word (e.g., several phonemes or syllables). Old adults display reduced speech fluency, reduced speech rate, increased pause duration, and stable/increased lexical diversity compared to younger adults [4] [9] [10] [11]. In connected speech or in discourse, old adults who have experienced word-finding difficulties in single word production tasks are endowed with more flexibility to navigate their vocabulary reservoir and thus display preserved or increased lexical diversity than younger counterparts [12]. For example, in picture description tasks, old adults are found to use more unique, non-normative nouns than younger adults [13].

At the sentence level, old adults display reduced syntactic complexity in spoken and written language, which is argued to be related to reduced working memory capacity [14] [15].

 10.21437/Interspeech.2024-1871

Nevertheless, older adults are better at switching between different syntactic alternatives and aligning their syntactic choices with others in dialogue [16]. At the discourse level, older adults produce more off-topic speech under some conditions [17]. A decline in density of ideas is found in written essays and spoken picture descriptions [14] [18].

The aforementioned literature suggests that age-related changes in language ability are far from conclusive. One explanation for the inconclusive findings is the co-existence of both preservation (mostly in conceptual representations) and decline (predominately in language production) in language ability as people age healthily. Second, most results about age-related language changes were gained from well-controlled experimental studies under a lab setting, such as in lexical retrieval tasks and picture description tasks [19] [20], which are different from the way we use language in a natural setting. Different language data types (e.g., scripted speech, semi-structured conversation, free speech) may return various results about language use in an aging process. Third, from a practical perspective, the manual assessments of language abilities in previous research are subjective, hence the results can be hard to be replicated. To overcome the limitations, we (i) focus on language production at the word level; (ii) using natural speech data; and (iii) applying objective quantitative measures to the speech data. Details are introduced below.

2.2. Quantifying Language with Objective Measures: Part-of-Speech (PoS) Distribution and Lexical Concreteness

The distribution of different PoS categories in the text is one of the most used lexical measures of language samples. While content words reflect the theme of the talk or speakers' attention, function words, though fewer in type than content words, reflect people's thinking styles [21]. Age differences in function words were found to be robust. For example, among participants age from 8 to 70 years old, older speakers use more exclusive words, more we, and fewer I than younger speakers do [21]. More recently, a fine-grained analysis of healthy people compares densities of different syntactic categories and other lexical characteristics between two age groups. Older people are found to use more verbs, pronouns, fewer prepositions, conjunctions, and determiners than younger speakers [20].

Lexical concreteness is a well-studied psycholinguistic parameter, measuring the degree to which the referent of a word can be experienced by senses or actions [22]. It has been consistently found that changes in lexical concreteness can be a marker for pathological aging. Patients with semantic dementia display more abstract lexical uses than healthy counterparts [23]. Nevertheless, the concreteness changes in healthy population are understudied with a few mixed findings. Some find an insignificant age effect on lexical concreteness; others find increased concreteness in older adults [20] [24]. In the context of these inconclusive reports, it is reasonable to revisit the concreteness features of normal aging population.

To our knowledge, few studies have investigated both lexical density and concreteness in free speech of healthy people aged above 60 years old. This paper reports a corpus-based study about the age-related differences in lexical density and concreteness in a normal aging population.

2.3. The Effect of Sex on Language Use

Since we aim at exploring lexical features in an aging process, age is the primary factor to be investigated in this study. Besides the effect of age, sex also affects language uses independently [11]. Due to physiological differences and sex-differentiated anatomical changes, men and women display distinct aging patterns [25]. It has been found that women do not show decrease in fluency across their life span, while men display increased speech disfluencies after 45 years old [11]. From the perspective of general communication purposes, men tend to use language primarily for conveying information, while women use it more for social purposes [26]. An analysis of 14,000 text samples shows that men referred more to object properties and impersonal topics, while women use more words related to psychological and social processes [27]. In discourse, women's language is often characterized by tentativeness and uncertainty, by more hedges, overly polite forms, and apologies in their talk than men [27] [28].

The sex-related differences in language are also closely associated with speakers' socioeconomical conditions [29]. In this case, socioeconomical variables can also be confounding variables in modeling language changes in aging, disguising the real effect of age. Given the multi-faced nature of aging languages, previous research has tried to incorporate age and sex as two primary factors accounting for language variances, supplemented with other demographic variables such as education and multilingual experiences that may affect the trajectory of cognitive aging [20] [30]. In this study, we only consider age and sex as the two main factors influencing language use. Given that the level of education and the number of languages spoken by participants are balanced, we do not covary these variables in our analysis.

3. Method

3.1. Participants and Materials

The subjects were participants in the project *Community Health and Intergenerational* (CHI) study, a cohort study of aging in community-dwelling adults aged 60 and older in Singapore [31]. Many Singaporeans of this age cohort are bilingual or multilingual, speaking English and one or more of the heritage languages of Chinese, Malay and Tamil. Participants are all cognitively healthy. Neuropsychological battery of tests was used for the diagnosis of neurocognitive disorders and normal aging [32]. Speakers aged above 80 years old were excluded due to sporadic age data points. Speakers with less than 13 years of education were excluded due to the incompletion of high-school education. After exclusion, there were 243 participants in total (133 women, 110 men). The demographic information of the participants is shown in Table 1.

Table 1: *Participants' demographic information. The units of age and education are in years.*

Variable		**Men**	**Women**	**Sig.**
Age	Mean	66.09	67.01	.079
	SD	5.08	4.67	
Education	Mean	16.38	16.67	.239
	SD	2.55	2.39	
Language	Mean	2.36	2.35	.853
	SD	0.56	0.55	

The age, sex, education, and the number of languages spoken by participants were uncorrelated with each other, as shown in Table 2.

Table 2: *The correlations between age, sex, education, and the number of languages spoken by participants.*

Variable		Age	Sex	Education	Language
Age	ρ	1	0.11	-0.08	0.09
	p	.	.079	.222	.183
Sex	ρ	0.11	1	0.08	0.01
	p	.079	.	.240	.853
Education	ρ	-0.08	0.08	1	-0.05
	p	.222	.240	.	.472
Language	ρ	0.09	0.01	-0.05	1
	p	.183	.853	.472	.

For every participant, we collected about 15-min free speech. Participants were asked to speak freely about their life and experiences in English with minimal involvement from the interviewers. Speeches were recorded with audio recorders in an ordinary office setting and transcribed into text files. Then the transcriptions were then tagged with PoS information with Stanford Parser version 4.2 [33]. We followed the Penn Treebank tag set with adjustments for Singapore English (see Table 3) and checked by trained research assistants [34] [35]. In total, we got 426,849 tagged word tokens.

Table 3: *The Penn Treebank Tag Set for Singapore English*

PoS Categories	Tags
Noun	NN (*book*), NNS (*books*), NNP (*Singapore*), NNPS (*Times*)
Verb	VB (*be*), VBD (*were*), VBG (*being*), VBN (*been*), VBP (*are*), VBZ (*is*)
Adjective	JJ (*good*), JJR (*better*), JJS (*best*)
Adverb	RB (*slowly*), RBR (more), RBS (*most*)
Modal verb	MD (*can*)
Pronoun	PRP (*he*), PRP$ (*his*)
Wh-word	WP (*what*), WP$ (*whose*), WRB (*when*), WDT (*which*)
Determiner	DT (*the*), PDT (*all*)
Preposition	IN (*at*)
Numeral	CD (*two*)
Conjunction	CC (*and*)
Existential *there*	EX (*there*)
Particle	RP (give *up*)
Infinitive *to*	TO (*to*)
Sentence final particle	SFP (*lah*)
Local use GOT	GOT (got)
Interjection	UH
Fragment	FRG
Symbol	SYM

3.2. Textual Data Processing

We reported the lexical features of the speech sample in two steps. First, we calculated several global lexical features to describe the speech sample, including the total number of all words, the total number of major content word categories, lexical diversity, speech time, and speech rate. Second, we reported results of regression analyses between age, sex, and target lexical measures.

For the calculation of global lexical measures, we summed the token and type count of all PoS categories, nouns, verbs, and adjectives for every subject. The count of adverbs was not calculated in the present study, since the adverb category is highly heterogeneous and hard to be generalized with a clear definition across languages [36]. Moreover, the adverb category in our corpus contains some interjective-like words, such as "so", "then", "well", etc, which makes the adverb category less representative as a content word class. Lexical diversity typically is measured with Moving-Average Type-Token Ratio (MaTTR), which is insensitive to text length. MaTTR calculates Type-Token Ratio (TTR) with a fixed length of window of tokens, moving one word at a time from the beginning to the end of a text, and averages the measured TTRs of all windows [37]. We set a window of 20 words for the present dataset. Speech time was measured by the duration time (in minutes) of the recordings. Speech rate was calculated by dividing the total number of word tokens by speech time, the unit is word count per minute.

For the associations between age and lexical features, we examined four lexical measures: speech rate, lexical diversity, lexical density, and lexical concreteness. Lexical densities of different PoS categories were calculated by dividing the token number of words from a PoS category over the total number of all word tokens, then multiplied by 100. We calculated the densities of major content words: nouns, verbs, adjectives; and major function words: pronouns, determiners, prepositions, modal verbs, and interjections.

Lexical concreteness is quantified by published norms, with 1 being the most abstract and 5 being the most concrete [38]. Since lexical concreteness is only meaningful for content words, we calculated the concreteness scores of nouns, verbs, and adjectives. The concreteness scores of adverbs were not calculated due to the ambiguous semantic value of the adverb category. Word inflections were not rated by the concreteness rating list in [38], we calculated inflected words with the score of the lemma (pass for passed, colleague for colleagues). After getting the score of each word, we multiplied it with the number of word tokens to get a word's total score. The sum of all word tokens was then divided by the total number of word tokens to arrive at the concreteness score of each subject's speech sample.

After getting the lexical data of every subject, we conducted linear regression to explore the effects of age and sex on language use with reference to adjusted R-squares. The dependent variables are lexical measures. The independent variables include one continuous variable, age, and one categorical variable, gender. We encoded gender with numeric values, with women as 1, and man as 2.

4. Results

Table 4 shows the textual properties of the dataset.

Table 4: *Descriptive statistics about the main lexical characteristics of the speech sample. The unit of talk time is in minute, speech rate in word per minute.*

Lexical Measures	SUM	Mean	SD
Talk Time	3,178	13.08	5.18
Speech Rate	-	137.62	25.61
All words, type	111,722	459.76	155.26
All words, token	426,849	1756.58	867.52
Nouns, type	37,006	152.29	63.39
Nouns, token	75,094	309.03	149.17
Verbs, type	32,369	133.21	50.94
Verbs, token	87,596	360.48	190.46
Adjectives, type	12,847	52.87	21.55
Adjectives, token	22,779	93.74	46.81

Table 5 shows the regression analysis of lexical features as a function of age and sex.

Table 5: *Regression for lexical characteristics as a function of age and sex.*

Lexical Measures	**Age**		**Sex**		**R^2**
	β_1	*p*	β_2	*p*	
Lexical Diversity	-0.01	.88	-0.20	**.00**	0.03
Speech Rate	-0.16	**.02**	-0.13	**.03**	0.05
Noun (concreteness)	0.18	**.01**	-0.23	**.00**	0.07
Verb (concreteness)	0.03	.61	0.05	.44	0.00
Adjective (concreteness)	0.17	**.01**	0.05	.46	0.03
Noun	0.09	.16	0.28	**.00**	0.09
Verb	-0.09	.14	-0.25	**.00**	0.07
Adjective	-0.01	.94	0.19	**.00**	0.03
Determiner	0.02	.71	0.30	**.00**	0.08
Preposition	-0.07	.25	0.18	**.01**	0.03
Interjection	-0.14	**.04**	0.10	.13	0.02
Pronoun	0.01	.88	-0.40	**.00**	0.15
Modal Verb	-0.08	.24	0.10	.12	0.01

According to Table 5, the effect of age is significant on speech rate (β_1= -0.16, p= .02), noun concreteness (β_1= 0.18, p= .01), adjective concreteness (β_1= 0.17, p= .01), and density of interjections (β_1= -0.14, p= .04). Older speakers are associated with a lower speech rate, more concrete nouns and adjectives, as well as a lower density of interjections.

The sex effect is significant on lexical diversity (β_2= -0.20, p< .01), speech rate (β_2= -0.13, p= .03), concreteness of nouns (β_2= -0.23, p< .01), densities of nouns (β_2= 0.28, p< .01), verbs (β_2= -0.25, p< .01), adjectives (β_2= 0.19, p< .01), determiners (β_2= 0.30, p< .01), prepositions (β_2= 0.18, p= .01), and pronouns (β_2= -0.40, p< .01). Women are associated with a faster speech pace, higher lexical diversity, higher densities of verbs and pronouns. Men are associated with higher densities of nouns, adjectives, prepositions, and determiners.

5. Discussion

5.1. The Age Effect

The positive association between age and lexical concreteness in our study are different from existing findings (e.g., [20]), which may be due in part to the limited range of concreteness that could be achieved in semi-structured speech tasks. Natural speech in our study, by contrast, provides valuable data for the assessment of cognitive abilities. Besides, we found age-related changes in concreteness differ across PoS categories. The dissociation between nouns and verbs are consistent with the findings reported in the linguistic and neuropsychological literatures [39] [40].

Previous studies have observed increased uses of pronouns, filler words, and lexical diversity in older speakers [4] [8]. In our corpus, however, we did not find a significant positive effect of age on lexical diversity, consistent with findings in [20]. One explanation for such a discrepancy between our finding and existing results is that the 20-year range of participants (i.e., 60-79) in the present study is much smaller than a 40-year range in other studies (e.g., people in their 20s vs. people in their 60s) so the age-related differences in our sample are too small to be detectable. Besides, we found the age effect on pronoun density is not significant, though the direction of correlation is still positive. The insignificant age effect on pronoun density may be due to the co-existence of a diminished size of social networks, which lead to reduced pronoun uses, and an increased focus of self, which is associated with increased pronoun use [26]. Finally, we found age has a significant negative effect on the density of filler words, opposite from existing findings [9]. Although our speech sample did not show an age-related increase in the density of fillers, we revealed a significant negative effect of age on speech rate. The slower speech rate among older speakers may have modified the decreasing trend of filler words.

5.2. The Sex Effect

The sex effect on lexical diversity is consistent with the results in that men generally display a lower lexical diversity than women [41]. Besides, men are associated with a slower speech rate, in line with the finding that men are more likely to display speech disfluencies than women at an old age [11]. For content word distribution, men's higher use of nouns and adjectives, and women's higher use of verbs reveal different communication purposes, consistent with the results in [42]. For function word distribution, our findings are consistent with the results in [29] and [42]: women use more pronouns, men use more prepositions and determiners. Possible reasons include women being generally more self-focused than men, and women having lower social status than men, leading to more use of pronouns [43]. Men's higher use of articles and prepositions, by contrast, characterizes a concrete thinking and a tendency of categorization [42].

6. Conclusions

This study shows that lexical characteristics in aging can be detected by corpus data and psycholinguistics metrics. Overall, lexical uses remain stable in an aging process. Lexical distribution is predominately affected by age ands sex, while lexical concreteness is more affected by age.

We demonstrated that by incorporating NLP tools and psycholinguistic metrics into the construction of natural language datasets, it is possible to model language variations in healthy aging and set a benchmark for cognitive health evaluation. The statistical results demonstrated that automatic analysis of speech data enables the development of a normative baseline of language variances in normal aging populations, which may be used to assess pathological aging through linguistic markers in the future.

7. Acknowledgements

I would like to thank Professor Zhiming Bao for giving invaluable suggestions during the writing of this paper.

8. References

[1] J. E. Peelle, "Language and aging," in *The Oxford Handbook of Neurolinguistics*, M. de Zubicaray and M. Schiller, Eds. Oxford University Press, 2019.

[2] C. F. M. Poulisse, "Language comprehension in healthy aging and mild cognitive impairment," PhD dissertation, University of Birmingham, 2020.

[3] M. Bilodeau-Mercure, C. L. Lortie, M. Sato, M. Guitton, and P. Tremblay, "The neurobiology of speech perception decline in aging," *Brain Structure and Function*, vol. 220, pp. 979-997, 2015.

[4] W. S. Horton, D. H. Spieler, and E. Shriberg, "A Corpus Analysis of Patterns of Age-Related Change in Conversational

Speech," *Psychology and Aging*, vol. 25, no. 3, pp. 708–713, 2010.
[5] C. M. Brown, P. Hagoort, and M. Kutas, "Postlexical integration processes during language comprehension: Evidence from brain-imaging research," in *The new cognitive neurosciences*, MIT Press, pp. 881-895, 2000.
[6] G. Waters and D. Caplan, "Age, working memory and on-line syntactic processing in sentence comprehension," *Psychology and Aging*, vol. 16, pp. 128–144, 2001.
[7] M. Kutas and K. D. Federmeier, "Electrophysiology reveals semantic memory use in language comprehension," *Trends in Cognitive Science*, vol. 4, pp. 463–470, 2001.
[8] E. A. L. Stine-Morrow, D. D. Gagne, D. G. Morrow, and B. H. DeWall, "Age differences in rereading," *Memory and Cognition*, vol. 32, pp. 696–710, 2004.
[9] H. Bortfeld, S. D. Leon, J. E. Bloom, M. F. Schober, and S. E. Brennan, "Disfluency rates in conversation: effects of age, relationship, topic, role, and gender," *Language and Speech*, vol. 44, no. 2, pp. 123–147, 2001.
[10] J. Bóna, "Temporal characteristics of speech: The effect of age and speech style," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. EL116–EL121, 2014.
[11] F. Moscoso del Prado Martín, "Vocabulary, Grammar, Sex, and Aging," *Cognitive Science*, vol. 41, no. 4, pp. 950–975, 2017.
[12] G. V. Pashek and C. A. Tompkins, "Context and word class influences on lexical retrieval in aphasia," *Aphasiology*, vol. 16, pp. 261–286, 2002.
[13] G. Kavé and S. Nussbaum, "Characteristics of noun retrieval in picture descriptions across the adult lifespan," *Aphasiology*, vol. 26, pp. 1238–1249, 2012.
[14] S. Kemper, M. Thompson, and J. Marquis, "Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content," *Psychology and Aging*, vol. 16, pp. 600–614, 2001.
[15] S. Kemper and A. Sumner, "The structure of verbal abilities in young and older adults," *Psychology and Aging*, vol. 16, no. 3, pp. 12–322, 2001.
[16] D. J. Davidson, R. T. Zacks, and F. Ferreira, "Age Preservation of the Syntactic Processor in Production," *Journal of Psycholinguistic Research*, vol. 32, no. 5, pp. 541–566, 2003.
[17] T. Y. Arbuckle and D. P. Pushkar Gold, "Aging, inhibition and verbosity," *Journal of Gerontology: Psychological Sciences*, vol. 48, pp. 225–232, 1993.
[18] C. Mackenzie, "Adult spoken discourse: The influences of age and education," *International Journal of Language and Communication Disorders*, vol. 35, pp. 269–285, 2000.
[19] M. Nicholas, L. Obler, M. Albert, and H. Goodglass, "Lexical retrieval in healthy aging," *Cortex*, vol. 21, no. 4, pp. 595–606, 1985.
[20] S. Cho et al., "Lexical and Acoustic Characteristics of Young and Older Healthy Adults," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 2, pp. 302–314, 2021.
[21] J. W. Pennebaker and L. D. Stone, "Words of wisdom: Language use over the life span," *Journal of Personality and Social Psychology*, vol. 85, pp. 291–301, 2003.
[22] A. Paivio, J. C. Yuille, and S. Madigan, "Concreteness, imagery, and meaningfulness values for 925 nouns," *Journal of Experimental Psychology*, vol. 76, no. 1p2, pp. 1–25, 1968.
[23] M. F. Bonner et al., "Reversal of the concreteness effect in semantic dementia," *Cognitive Neuropsychology*, vol. 26, no. 6, pp. 568–579, 2009.
[24] H.-W. Huang, A. M. Meyer, and K. D. Federmeier, "A "concrete view" of aging: Event related potentials reveal age-related changes in basic integrative processes in language," *Neuropsychologia*, vol. 50, no. 1, pp. 26–35, 2012.
[25] E. Nakamura and K. Miyao, "Sex differences in human biological aging," *Journal of Gerontology A: Biological Sciences*, vol. 63, pp. 936–944, 2008.
[26] A. Colley et al., "Style and content in emails and letters to male and female friends," *Journal of Language and Social Psychology*, vol. 23, pp. 369–378, 2004.
[27] M. L. Newman et al., "Gender Differences in Language Use: An Analysis of 14,000 Text Samples," *Discourse Processes*, vol. 45, no. 3, pp. 211–236, 2008.
[28] R. Lakoff, "Language and woman's place," *Language in Society*, vol. 2, no. 1, pp. 45–80, 1973.
[29] A. Mulac, L. L. Torborg, and J. J. James, "Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect," *Communication Monographs*, vol. 53, pp. 115–129, 1986.
[30] S. E. Nijmeijer et al., "Musical and Multilingual Experience Are Related to Healthy Aging: Better Some Than None But Even Better Together," *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, vol. 78, no. 4, pp. 609–619, 2023.
[31] R. Z. Y. Lee et al., "CHI study: protocol for an observational cohort study on aging and mental health in community-dwelling older adults," *BMJ Open*, vol. 10, no. 5, e035003, 2020.
[32] M. L. Lim et al., "Cross-cultural application of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): performances of elderly Chinese Singaporeans," *The Clinical Neuropsychologist*, vol. 24, no. 5, pp. 811-826, 2010.
[33] D. Klein and C. D. Manning, "Accurate Unlexicalized Parsing," in *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430, 2003.
[34] M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz, "Building a large annotated corpus of English: The Penn Treebank," *Computational Linguistics*, vol. 19, no. 2, pp. 313–330, 1993.
[35] L. Lin et al., "Tagging Singapore English," *World Englishes*, vol. 42, no. 4, pp. 624–664, 2023.
[36] R. Quirk et al., A Comprehensive Grammar of the English Language. New York: Longman, 1985.
[37] M. A. Covington and J. D. McFall, "Cutting the Gordian knot: The moving-average type–token ratio (MATTR)," *Journal of Quantitative Linguistics*, vol. 17, no. 2, pp. 94–100, 2010.
[38] M. Brysbaert, A. B. Warriner, and V. Kuperman, "Concreteness ratings for 40 thousand generally known English word lemmas," *Behavior Research Methods*, vol. 46, no. 3, pp. 904–911, 2014.
[39] A. Caramazza and A. E. Hillis, "Lexical organization of nouns and verbs in the brain," *Nature*, vol. 349, pp. 788-790, 1991.
[40] G. Vigliocco et al., "Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies," *Neurosience and Biobehavioral Reviews*, vol. 35, pp. 407-426, 2011.
[41] J. Piersoul and F. Van de Velde, "Men use more complex language than women, but the difference has decreased over time: a study on 120 years of written Dutch," *Linguistics*, vol. 61, no. 3, pp. 725-747, 2023.
[42] C. K. Chung and J. W. Pennebaker, "The psychological functions of function words," in *Social Communication*, K. Fiedler, Ed. New York: Psychology Press, pp. 43-359, 2007.
[43] M. L. Newman et al., "Lying words: Predicting deception from linguistic style," *Personality and Social Psychology Bulletin*, vol. 29, pp. 665–675, 2003.