Original Article | | Peer Reviewed

USING CORPUS ANALYSIS TO INVESTIGATE LINGUISTIC DISORDERS IN SPONTANEOUS SPEECH

Journals of Arts & Humanities Studies
ISSN: 3069-325X (Online)
Vol. 1: Issue 4
Page 11–19 © The Author(s) 2025
Received: 30 September 2025
Accepted: 21 October 2025
Published: 23 November 2025

Mai Ahmed - *Language Convergence Research Center, Chosun University, Gwangju, South Korea; Department of Korean Language and Literature, Ain-Shams University, Cairo, Egypt*

Soon-Bok Kwon - *Department of Language and Information, Pusan National University, Busan, South Korea*

ABSTRACT

This study investigates the application of corpus analysis techniques for the automated assessment of linguistic features in Korean speech, specifically within the context of language disorders associated with cognitive decline. Addressing the limitations of traditional, manual linguistic analysis, this paper proposes a methodology utilizing the corpus morpheme analysis and tagging to quantify semantic and syntactic variables indicative of language disorders. The method details the use of part-of-speech tagging to derive indices related to lexical diversity, semantic coherence, and syntactic complexity, accounting for the unique grammatical characteristics of the Korean language. By automating the analysis of linguistic biomarkers, this approach aims to enhance the precision and efficiency of analyzing and monitoring language disorders in cognitive decline and other language-related pathologies in Korean speakers. While acknowledging the challenges posed by the complexity of Korean morphology, this study concludes that integrating corpus analysis into language pathology research holds significant potential for advancing diagnostic and research methodologies. Future work should focus on validating this approach across various language disorders and developing specialized corpus analysis tools tailored for Korean language pathology assessments.

KEY WORDS

corpus analysis, cognitive decline, language disorders, speech analysis

Introduction

Neural pathway damage in the brain, whether due to age-related degeneration or disease, negatively impacts speech production and comprehension. For instance, both aging and neurological conditions can impair the brain's memory system (Bak et al., 2022; Tae et al., 2025). Dementia, a cognitive disorder characterized by memory decline, leads to significant difficulties in word retrieval and speech production (Bäckman et al., 2001; Chen et al., 2001; Taler & Philips, 2008). Individuals with Alzheimer's disease often exhibit impaired recall, resulting in hesitations and word-finding difficulties during conversations. This observation underscores the importance of incorporating linguistic assessments into cognitive evaluations, thereby facilitating the measurement of cognitive dysfunction through language analysis.

Research has consistently demonstrated the strong influence of semantic and lexical variables on language production in the early stages of Alzheimer's disease (Vigo et al., 2022). Taler and Philips (2008) further established that language skills, including semantic content and lexical fluency, can be compromised in the initial phases of cognitive impairment. Moreover, studies have shown a correlation between the

progression of cognitive decline in Alzheimer's patients and a reduction in sentence comprehension, indicating that cognitive decline affects both receptive and expressive language abilities (Croot et al., 1999; Gumus et al., 2024).

Fraser et al. (2019) demonstrated that linguistic features related to semantic processing, such as word frequency and content density, effectively identify cognitive decline, particularly in its early stages. Similarly, Yeung et al. (2021) identified a decline in cohesion, coupled with word-finding difficulties and repetitive language use, as crucial markers for cognitive impairment. Bueno-Cayo et al. (2022) explored the correlation between lexical density and the degree of cognitive impairment and found that higher lexical density corresponded to better cognitive function. While Forbes-McKay and Venneri (2005) found that individuals with Alzheimer's disease demonstrated a heightened tendency to use indefinite terms, reflecting a decline in semantic processing abilities associated with cognitive decline. This observation is further supported by Gumus et al. (2024), who reported a decrease in noun usage and an increase in pronoun usage with declining cognitive function. This trend aligns with findings from Mueller et al. (2018), Fraser et al. (2016, 2019), and Pompili et al. (2020), all of which suggest that patients with cognitive decline favor common terms like pronouns over nouns and verbs, exhibit a preference for shorter words, and display reduced lexical variety. On the other hand, Kemper et al. (1993) argued that Alzheimer's patients showed a decrease in sentence length, particularly in the number of clauses, indicating challenges in constructing complex sentences and maintaining logical connections. Gumus et al. (2024) also reported that patients with neurodegenerative diseases tend to use simpler sentence structures, shorter vocabularies, and fewer prepositional phrases.

Recent investigations into linguistic changes associated with neurological diseases have focused on automating the analysis of these linguistic features. Traditional methods, such as manual calculation of mean length of utterance (MLU) and T-unit differentiation, are time-consuming and inefficient for large-scale, precise analyses. In an era of automation, there is a pressing need for optimized, automated analysis methods, highlighting the significance of corpus data analysis. Studies by Gumus et al. (2024), Abiven and Ratté (2020), and Huang et al. (2024) have utilized part-of-speech (POS) tagging to quantify semantic features, such as lexical diversity and noun-verb ratios, to distinguish between patient and non-patient speech. These studies underscore the potential of automated speech feature analysis for investigating linguistic disorders related to cognitive decline.

However, the utilization of POS tagging for analyzing speech features in Korean-speaking patients with language pathologies appears to be limited. While POS tagging is employed in various linguistic and natural language processing tasks for Korean, its application in clinical settings for assessing language disorders in Korean speakers is not well documented. This might be due to the difficulties the development of a POS tagging tool for Korean has faced. Kim et al. (2024) highlight that some Korean text analysis tools are still predicated on adaptations of English text analysis applications (Gho & Lee, 2020). This approach can lead to inaccuracies in morpheme separation, often resulting in word conjugation errors. This issue stems from fundamental differences in word and sentence formation between Korean and English. Korean relies on suffixes and grammatical particles, whereas English lacks such grammatical elements. Nonetheless, significant efforts are dedicated to refining Korean text analysis at the morpheme level, focusing on improving accuracy within the framework of Korean grammatical rules (Kim & Choi, 2018; Lee et al., 2018; Park & Cho, 2014; Shin & Ock, 2012).

This paper aims to investigate using POS tagging to facilitate analyzing speech linguistic characteristics in Korean speech, thereby expanding the analytical tools available for research.

Method

Participants

This study utilized a corpus of spontaneous speech samples collected from 39 participants residing in the Busan region, with a mean age of 81.9 years. To investigate the relationship between cognitive ability and speech characteristics, participants' spontaneous speech data were systematically categorized into three distinct groups based on their scores from the Cognitive Impairment Screening Test (CIST). The CIST scores served as a crucial indicator of the participants' cognitive health. Comprehensive demographic information for all participants is presented in Table 1.

Table 1. Demographic information by group

Category	NC	MCI	De	*P-value*
N (%)	13(33.3%)	13(33.3%)	13(33.3%)	-
Age (*SD*)	81.4(5.68)	83.4(6.92)	80.8(6.95)	0.572
Education (*SD*)	7.46(4.25)	6.9(2.66)	7.9(5.8)	0.847
CIST (*SD*)	21.6(4.09)	12(3.24)	7.23(4.06)	<.001***

****p*<.001 Note. De: dementia; MCI: mild cognitive impairment; NC: normal cognitive; *SD*: standard deviation.

Linguistic feature analysis using tag sets

This study is utilizing UTagger-4 for morpheme tagging of the collected speech samples after transcription. Utagger-4 was specifically chosen due to its robust training on the extensive Sejong corpus and Korean dictionaries, which significantly enhances its performance for morpheme registration. This tool boasts a high analysis accuracy, reportedly reaching up to 95%.[1] Our selection of Utagger-4 was further motivated by its user-friendly web-based interface, which streamlined the analysis process. The absence of a coding requirement allowed for direct text input and analysis via a simple click, ensuring accessibility and efficiency. Following the morpheme tagging, we employed AntConc to quantify the linguistic features by systematically searching and recording the frequencies of the tags and morphemes within our corpus. This quantitative data was then organized in an Excel sheet to facilitate the application of relevant linguistic equations and subsequent feature analysis.

The following table presents the Sejong Tag Set[2], which will be utilized in the development of equations for analyzing linguistic characteristics of Korean language disorders.

Table 2. Tag set of Korean morphemes

POS		Tag	POS	Tag
Substantive	Noun	NN	Common Noun	NNG
			Proper Noun	NNP
			Dependent Noun	NNB
	Pronoun	NP	Pronoun	NP
	Numeral	NR	Numeral	NR
Predicate	Verb	VV	Verb	VV
	Descriptive Verbs	VA	Descriptive Verbs	VA
	Auxiliary	VX	Auxiliary	VX
Modifiers	Adjective	MM	Adjective	MM
	Adverb	MA	General adverbs	MAG
			Conjunctions	MAJ
Independent words	Exclamation	IC	Exclamation	IC
Dependent forms	Endings	E	Pre-final ending	EP
			Sentence-closing	EF
			Connective	EC
			Nominal transformative	ETN
			Adnominal transformative	ETM
Postposition	Case-marker	JK	Subject	JKS
			Complement	JKC
			Adnominal	JKG
			Object	JKO
			Adverbial	JKB
			Vocative	JKV
			Citation	JKQ
	Auxiliary Postpositional	JX	Auxiliary Postpositional	JX
	Conjunctive Postpositional	JC	Conjunctive Postpositional	JC
Symbols	Full stop, question mark, exclamation mark			SF

(As adapted from: Sejong tagset (2008): https://jchern96.tistory.com/12375684)

[1] http://klplab.ulsan.ac.kr/doku.php?id=utagger_4

[2] The Sejong Tag Set, developed as part of the Korean government's Sejong Project in the early 2000s, is designed to reflect Korean linguistic theory (National Institute of the Korean Language, 2003).

The application of the Sejong Tag Set (Table 1) enables the derivation of significant information regarding semantic and syntactic disorders in Korean speech. Specifically, Coherence indices can be derived by analyzing the ratio of information word units to functional word units. Information units, such as nouns and verbs, contribute semantic content, whereas functional units primarily serve grammatical roles.

Information word ratio (IWR):

$$IWR = \frac{\text{NNG} + \text{NNP} + \text{NP} + \text{VV} + \text{VA} + \text{MMD} + \text{MMN} + \text{MMA} + \text{MAG}}{T}$$

T: total number of words in text

Functional word ratio (FWR):

$$FWR = \frac{\text{NNB} + \text{IC} + \text{VX}}{T}$$

T: total number of words in text

Lexical diversity can be quantified through type-token ratio. Morphological-level analysis enhances word type count accuracy by segmenting eojeols into individual morphemes, thereby examining word formation at the stem level.

Type-to-token ratio (TTR):

$$TTR = \frac{\text{V}}{T}$$

V: total number of words by word type
T: total number of words in text

Syntactic feature analysis is also feasible due to the rich grammatical structure of Korean. For example, sentence counts can be obtained from the final ending tag (EF) count. Clause counts can be derived from connective (EC) and transformative (ETM, ETN) ending counts. Mean length of utterance (MLU) can be estimated by dividing the total token count by the sentence-final tag (SF) count.

Mean Length of Utterance (MLU):

$$MLU = \frac{\text{T}}{SF}$$

T: total number of words (tokens) in text

Mean length of T-unit (MLTU):

$$MLTU = \frac{\text{EC+ETM+ETN}}{EF}$$

Mean length of AS-unit (MLAS):

$$MLAS = \frac{\text{EC+ETM+ETN}}{SF}$$

Incomplete-to-complete sentence ratio (ICS):

$$ICS = \frac{\text{SF}-\text{EF}}{EF}$$

Grammatical complexity (GC):

$$GC = \frac{\text{EP+EF+EC+ETN+ETM+JKS+JKC+JKG+JKO+JKB+JKV+JKQ+JX+JC}}{T}$$

Results

Table 3 presents the results of the one-way ANOVA analysis for the linguistic features within the data sample. The findings consistently indicate that cognitive decline more significantly impacts syntactic complexity-related features than lexical features. Specifically, both the information word ratio (IWR) and functional word ratio (FWR) showed no statistically significant differences across the groups. This lack of significant difference is visually supported by the boxplot in Figure 1, where the means of the three groups, particularly for information word ratio, are highly similar. While the functional word ratio did not reach

statistical significance, Figure 1 does illustrate a slight mean difference between the Mild Cognitive Impairment (MCI) and Normal Cognition (NC) groups, with the MCI group exhibiting a tendency to use more functional words. This observation aligns with previous literature arguing that certain functional words, such as pronouns and exclamations (fillers), are closely associated with cognitive decline. Furthermore, both type-to-token ratio (TTR) and content density also lacked significant inter-group differences, a pattern corroborated by Figure 1, suggesting that lexical diversity is not substantially affected by cognitive decline.

Conversely, syntactic factors reveal a different pattern. Although the p-value for Mean Length of Utterance (MLU) did not meet the conventional threshold for statistical significance (p=0.06), Figure 1 clearly depicts a notable difference between the Dementia (De) and NC groups, with the NC group demonstrating a higher mean length of words per utterance compared to the De group. Similarly, the Mean Length of T-unit (MLTU) also did not show a statistically significant difference across groups (p=0.22).

Crucially, the only statistically significant difference observed in the data was for the Mean Length of AS-unit (MLAS) across groups (p=0.03)*, strongly indicating its direct relevance to cognitive decline. This finding suggests that cognitive decline directly affects syntactic complexity per utterance or 'idea unit'. To elaborate, spontaneous speech, especially in individuals with cognitive decline, often features incomplete sentences, hesitations, and disjointed thoughts that might not form grammatically "complete" sentences but still convey an idea unit. While MLU, using the total words in text divided by all sentence-final tags (SF), is a broad measure of words per utterance, its inclusiveness of highly varied utterance types might obscure subtler changes. MLTU, relying on sentence-closing endings (EF) to identify grammatically complete sentences, may become less reliable if individuals with cognitive decline use fewer complete sentences or struggle to form them, thus reducing its sensitivity for capturing overall syntactic complexity across all spoken units. In contrast, MLAS utilizes all sentence-final tags (SF) as its denominator for clause-like units, effectively capturing the average length of discoursal or idea units, regardless of their grammatical completeness. This makes MLAS more robust for analyzing spontaneous speech where speakers might frequently use incomplete sentences or sentence fragments. Therefore, MLAS, by measuring clauses per idea unit, more acutely reflects syntactic simplification than MLU or MLTU, as it directly demonstrates the difficulty in constructing complex thoughts and sentences. While the NC group can produce complex syntactic structures, their spontaneous speech naturally incorporates more "incomplete" but functionally meaningful units. Conversely, the significant reduction in MLAS for the De group specifically highlights a decline in their ability to produce even these shorter, clause-like units efficiently.

Figure 1 further reinforces this by illustrating a reduction in clause-per-utterance usage as cognitive ability declines, collectively pointing towards a process of syntactic simplification with advancing cognitive impairment. However, calculating syntactic complexity by averaging the number of clauses per whole sentence did not yield any statistically significant group differences. This outcome may be attributable to the varying proportions of complete and incomplete sentence usage. As depicted in Figure 1, while the means for the incomplete-to-complete sentence ratio (ICS ratio) are highly similar across the three groups, the NC group exhibits a wider range. This wider range in the NC group, along with their higher frequency of incomplete sentence usage, could suggest greater cognitive flexibility, enabling transitions between topics and the incorporation of parenthetical sentences or interjections mid-utterance, which is indicative of more complex, spontaneous speech compared to the simplified syntax often observed with cognitive decline.

Table 3. Results of one-way ANOVA analysis of data

Variable	Mean	SD	F	F-crit	P-Value
IWR	0.90	0.05	0.53	3.26	0.59
FWR	0.18	0.08	2.25	3.26	0.12
TTR	0.66	0.10	0.40	3.26	0.67
MLU	5.15	2.06	3.13	3.26	0.06
MLAS	1.77	1.01	3.98	3.26	0.03*
MLTU	2.62	1.68	1.59	3.26	0.22
ICS ratio	0.55	0.76	0.81	3.26	0.45
GC	0.74	0.12	1.07	3.26	0.35

*$p < 0.05$, ICS: incomplete-to-complete sentence ratio, MLAS: Mean length of AS-unit, MLTU: Mean length of T-unit, MLU: Mean length of utterance, TTR: Type-to-token ratio, SD: Standard deviation, GC: Grammatical complexity

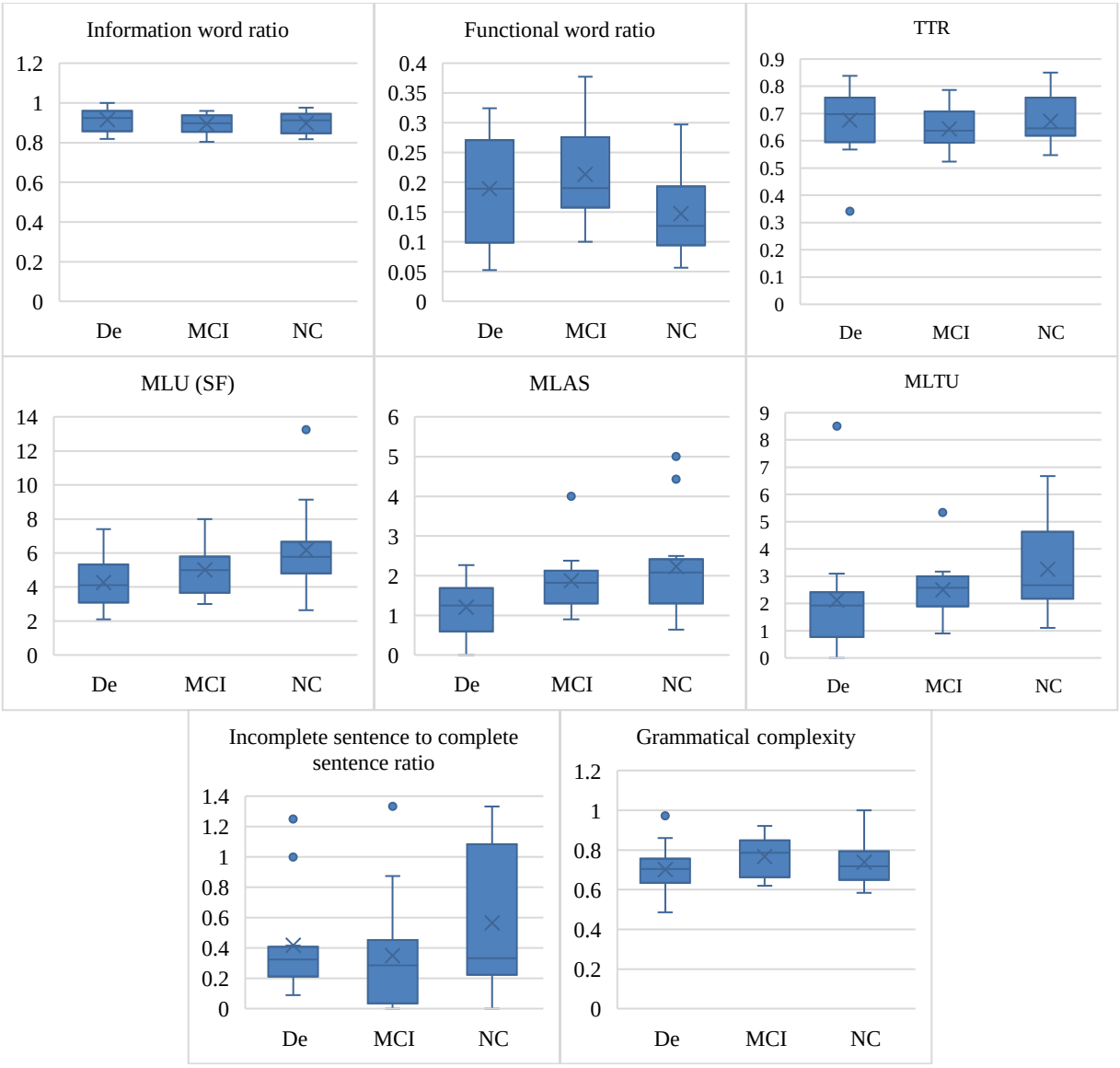

Figure 1. Boxplot of linguistic variables

Conclusion

This study explored the feasibility of employing corpus analysis techniques to analyze speech-language disorders, bridging two distinct fields to enhance research and analytical outcomes. By demonstrating the application of Korean text morphological analysis and its associated tag sets, we have illustrated how diverse linguistic features of Korean speech can be extracted. This approach aims to automate the analysis of Korean speech in language pathology research, thereby improving the precision and applicability of linguistic biomarkers for monitoring and detecting health conditions such as cognitive decline and dementia.

Our results consistently demonstrate a close relationship between syntactic complexity simplification and cognitive decline. This finding aligns robustly with previous research by Gumus et al. (2024) and Yeung et al. (2021) who also reported that cognitive impairment is reflected in the simplification of syntactic composition. Furthermore, Robin et al. (2021) observed that cognitively impaired individuals tend to exhibit slower speech and increased speech disfluency. This broader pattern of speech production changes provides a compelling explanation for the decrease in Mean Length of Utterance (MLU) identified in our findings, even if it did not reach statistical significance, reflecting a general reduction in verbal output length. More critically, the statistical significance of Mean Length of AS-unit (MLAS) underscores its specificity as a linguistic biomarker. MLAS, by capturing the average length of idea or clause-like units within spontaneous speech,

effectively reflects the reduced ability to construct complex internal syntactic structures. This makes it a particularly sensitive measure for detecting the subtle syntactic simplification characteristic of cognitive decline, distinguishing it from broader measures like MLU or MLTU which may be less attuned to the nuances of natural, often fragmented, conversational discourse.

Regarding lexical diversity-related features, our findings indicate that cognitive decline leads to an increased use of pronouns and fillers in spontaneous speech. This observation is extensively supported by several previous studies, including Gumus et al. (2024), Mueller et al. (2018), Fraser et al. (2016, 2019), and Pompili et al. (2020), all of which reported a preference for pronouns over proper nouns and verbs among subjects with cognitive impairment. This phenomenon is largely attributable to the challenges cognitive impairment patients face with object naming and word retrieval (Choi et al., 2013; Ha et al., 2009), which often prompt them to employ substitute words like pronouns or exclamations to compensate for lexical gaps in their speech. Specifically, an increase in fillers and exclamations with cognitive decline has been consistently reported (Mueller et al., 2018; Robin et al., 2021; Antonsson et al., 2021), directly linked to the speech fluency problems experienced during word retrieval difficulties.

Nonetheless, we posit that the integration of corpus analysis into language pathology research holds significant potential for advancing analytical methodologies and fostering novel research directions. The application of POS tagging for semantic content analysis extends beyond cognitive speech analysis, offering benefits for monitoring progress in other language disorders, including child language development and aphasia. For instance, POS tagging can aid in tracking MLU, T-unit, and AS-unit in children's language development, and in assessing semantic deficits in aphasia, which is closely linked to naming and word-finding difficulties.

Future research should prioritize evaluating the efficacy of corpus analysis across diverse language pathology domains. Comparative studies with established assessment methods are crucial to definitively determine its added value and optimize its clinical utility. Furthermore, the development of specialized corpus analysis tools tailored for the measurement of language disorder indicators across different language pathologies is strongly recommended.

References

Abiven, F., & Ratté, S. (2021). Multilingual Automation of Transcript Preprocessing in Alzheimer's Disease Detection. *Alzheimer's Dement., 7*, e12147. DOI: 10.1002/trc2.12147

Anthony, L. (2024). AntConc (Version 4.0.0) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Antonsson, M., Lundholm Fors, K., Eckerström, M., & Kokkinakis, D. (2021). Using a discourse task to explore semantic ability in persons with cognitive impairment. *Front. Aging Neurosci. 12*, 607449. DOI: 10.3389/fnagi.2020.607449

Bäckman, L., Small, BJ., & Fratiglioni, L. (2001). Stability of the Preclinical Episodic Memory Deficit in Alzheimer's Disease. *Brain, 124*, 96–102. DOI: 10.1093/brain/124.1.96

Bak, Y., Yu, S., Nah, Y., & Han, S. (2022). Bibliometric analysis of source memory in human episodic memory research. *Korean Journal of Cognitive Science, 33*(1), 23-50. DOI: 10.19066/cogsci.2022.33.1.002

Bueno-Cayo, AM., Del Rio Carmona, M., Castell-Enguix, R., Iborra-Marmolejo, I., Murphy, M., Irigaray, TQ., Cervera, JF., & Moret-Tatay, C. (2022). Predicting Scores on the Mini-Mental State Examination (MMSE) from Spontaneous Speech. *Behavioral sciences (Basel, Switzerland), 12*(9), 339. DOI: 10.3390/bs12090339

Chen, P., Ratcliff, R., Belle, SH., Cauley, JA., DeKosky, ST., & Ganguli, M. (2001). Patterns of Cognitive Decline in Pre-symptomatic Alzheimer's Disease: A Prospective Community Study. *Archives of General Psychiatry, 58*, 853–858. DOI: 10.1001/archpsyc.58.9.853

Choi, E., Sung, JE., Jeong, JH., & Kwag, E. (2013). Noun-Verb Dissociation in a Confrontation Naming Task for Persons with Mild Cognitive Impairment. *Dementia and Neurocognitive Disorders, 12*, 41-46. DOI: 10.12779/dnd.2013.12.2.41

Croot, K., Hodges, JR., & Patterson, K. (1999). Evidence for Impaired Sentence Comprehension in Early Alzheimer's Disease. *Journal of the International Neuropsychological Society, 5*, 393–404. DOI: 10.1017/S1355617799555021

Forbes-McKay, KE., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *26*(4), 243–254. DOI: 10.1007/s10072-005-0467-9

Fraser, KC., Meltzer, JA., & Rudzicz, F. (2016). Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's disease: JAD*, *49*(2), 407–422. DOI: 10.3233/JAD-150520

Fraser, KC., Lundholm Fors, K., Eckerström, M., Öhman, F., & Kokkinakis, D. (2019). Predicting MCI Status From Multimodal Language Data Using Cascaded Classifiers. *Frontiers in aging neuroscience*, *11*, 205. DOI: 10.3389/fnagi.2019.00205

Gho, Y., & Lee, G. (2020). Development of the Korean language text analysis program (KreaD index). *J Read Res, 56*, 225–45. DOI: 10. 17095/JRR.2020.56.8.

Gumus, M., Koo, M., Studzinski, CM., Bhan, A., Robin, J., & Black, SE. (2024). Linguistic Changes in Neurodegenerative Diseases Relate to Clinical Symptoms. *Front. Neurol., 15*, 1373341. Dol: 10.3389/fneur.2024.1373341

Ha, JW., Jung, YH., & Sim, HS. (2009). The Functional Characteristics of Fillers in the Utterances of Dementia of Alzheimer's Type, Questionable Dementia, and Normal Elders. *Korean Journal of Communication Disorders, 14*, 514-530.

Huang, L., Yang, H., Che, Y., & Yang, J. (2024). Automatic Speech Analysis for Detecting Cognitive Decline of Older Adults. *Front. Public Health, 12*, 1417966. Dol: 10.3389/fpubh.2024.1417966

Kemper, S., LaBarge, E., Ferraro, R. F., Cheung, H., Cheung, H., & Storandt, M. (1993). On the Preservation of Syntax in Alzheimer's Disease. *Archives of Neurology, 50*, 81–86. Dol: 10.1001/archneur.1993.00540010075021

Kim, SW., & Choi, SP. (2018). Research on joint models for Korean word spacing and POS (part-of-speech) tagging based on bidirectional LSTM-CRF. *J KIISE, 45*(8), 792–800. DOI: 10.5626/JOK.2018.45.8.792

Kim, DH., Ahn, S., Lee, E., & Seo, YD. (2024). Morpheme-based Korean Text Cohesion Analyzer. *SoftwareX, 26*, 101659. DOI: 10.1016/j.softx.2024.101659

Lee, Y., Noh, J., Song, M., & Shin, Y. (2018). An Improvement the accuracy of POS tagging for Korean Using CNN-LSTM. *The Korean Institute of Information Scientists and Engineers Conference Proceedings, 12*, 689–69.

Mueller, K., Koscik, R., Hermann, B., Johnson, S., & Turkstra, L. (2018). Declines in Connected Language Are Associated with Very Early Mild Cognitive Impairment: Results from the Wisconsin Registry for Alzheimer's Prevention. *Frontiers in aging neuroscience*, *9*, 437. Dol: 10.3389/fnagi.2017.00437

National Institute of Korean Language. (2003). *A Study on Korean Language Informatization for the 21st Century: Construction of the Sejong Corpus*. National Institute of Korean Language.

Park, E. & Cho, S. (2014). KoNLPy: Korean Natural Language Processing in Python". *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology, 2014*, 133-136.

Pompili, A., Abad, A., Martins de Matos, D., & Martins, IP. (2020). Pragmatic Aspects of Discourse Production for the Automatic Identification of Alzheimer's Disease. *IEEE Journal of Selected Topics in Signal Processing, 14*(2), 261-271. Dol: 10.1109/JSTSP.2020.2967879

Robin, J., Xu, M., Kaufman, LD., & Simpson, W. 2021. Using digital speech assessments to detect early signs of cognitive impairment. *Frontier Digital Health, 3*, 749758. DOI: 10.3389/fdgth.2021.749758

Shin, JC., & Ock, CY. (2012). A Korean Morphological Analyzer using a Pre-analyzed Partial Word-phrase Dictionary. *Journal of KIISE, JOK, 39*(5), 415-424. UCI: I410-ECN-0101-2013-569-002864889

Tae, J., Lee, Y., & Choi, W. (2025). The relationship between aging and inhibition ability: Evidence from a web-based Number Stroop task. *Korean Journal of Cognitive Science, 36*(1), 1-20. DOI: 10.19066/cogsci.2025.36.1.001.

Taler, V., & Phillips, NA. (2008). Language Performance in Alzheimer's Disease and Mild Cognitive Impairment: A Comparative Review. *J Clin Exp Neuropsychol, 30*, 501-556. DOI: 10.1080/13803390701550128

Utagger: http://klplab.ulsan.ac.kr/doku.php?id=utagger_4

Vigo, I., Coelho, L., & Reis, S. (2022). Speech- and Language-Based Classification of Alzheimer's Disease: A Systematic Review. *Bioengineering, 9*, 27. DOI: 10.3390/bioengineering9010027

Yeung, A., Iaboni, A., Rochon, E., Lavoie, M., Santiago, C., Yancheva, M., Novikova, J., Xu, M., Robin, J., Kaufman, LD., & Mostafa, F. (2021). Correlating Natural Language Processing and Automated Speech Analysis with Clinician Assessment to Quantify Speech-Language Changes in Mild Cognitive Impairment and Alzheimer's Dementia. *Alzheimer's research & therapy*, *13*(1), 109. DOI: 10.1186/s13195-021-00848-x.