Detecting Japanese Patients with Alzheimer's Disease based on Word Category Frequencies

Daisaku SHIBATA, Shoko WAKAMIYA, Eiji ARAMAKI
Nara Institute of Science and Technology (NAIST)
8916–5 Takayama, Ikoma City, 630–0192, Japan
{shibatadaisaku.rr8,wakamiya,aramaki}@is.naist.jp

Ayae KINOSHITA
Kyoto University Graduate School of Medicine
Syogoin, Sakyo-ku, Kyoto City, 606–8507, Japan
kinoshita.ayae.6v@kyoto-u.ac.jp

Abstract

In recent years, detecting Alzheimer's disease (AD) in early stages based on natural language processing (NLP) has drawn much attention. To date, vocabulary size, grammatical complexity, and fluency have been studied using NLP metrics. However, the content analysis of AD narratives is still unreachable for NLP. This study investigates features of the words that AD patients use in their spoken language. After recruiting 18 examinees of 53–90 years old (mean: 76.89), they were divided into two groups based on Mini Mental State Examination (MMSE) scores. The AD group comprised 9 examinees with scores of 21 or lower. The healthy control group comprised 9 examinees with scores of 22 or higher. Linguistic Inquiry and Word Count (LIWC). The word frequency was found from observation. Significant differences were confirmed for the usage of impersonal pronouns in the AD group. This result demonstrated the basic feasibility of the proposed NLP-based Alzheimer's disease detection approach.

1 Introduction

The increasing life expectancy has led to severe health and social problems. Among them, the most severe problem is the rising frequency of Alzheimer's disease (AD; Wortmann 2015) among the population. Japan is especially faced with a crisis posed by AD. Japan's Ministry of Health, Labour and Welfare reported that more than 1 in 4 control individuals would soon be afflicted with mild cognitive impairment (MCI) or AD. If all afflicted people were treated for MCI and AD, then the cost is estimated to be as high as 10 trillion dollars per year. As the number of patients with AD increases, the needs of these individuals might eventually exceed the current capacity of the national healthcare system, requiring various methods to detect the early stages of AD, prevent further deterioration, and alleviate requirements for care. Natural language processing (NLP) has also drawn much attention as a novel and simple method to detect AD using language.

Roark et al. indicated that a spoken narrative recall task can discriminate between healthy control people and those with MCI (Roark, Mitchell et al. 2011). Tanaka et al. proposed a computer avatar based approach to detect dementia in very early stages (Tanaka et al. 2016). Aramaki et al. also specifically examined the patients' narratives during a test (Aramaki et al. 2016). Year by year, MCI and AD narratives have been newly analyzed using NLP.

Although the details of the methods differ among them, they share the same approach, examining functional features (such as audio and transcript of narrative recall task by Roark et al.), spoken dialog by Tanaka et al. and transcriptions written and spoken narratives and vocabulary size by Aramaki et

al.) (Aramaki et al. 2016)) to detect and characterize patients with a disease. They did not deal with the contents of the narratives. In contrast, this paper presents a new method to detect AD based the categories of words patients use in spoken narratives. The word categories are classified by Linguistic Inquiry and Word Count (LIWC), a dictionary for text analysis.

To date, it has generally been pointed out that words from people with AD differ from those of healthy control people, including slowed speech, word-finding hesitation, sentences with abnormal words, and using words that are mispronounced or incomprehensible[1]. Especially, it is often said that AD patients more frequently use pronouns (e.g. it, that) than healthy control people. An example is presented in Figure 1. The sample includes much silence, repeating similar utterances, and pronouns.

Using the LIWC, this study empirically investigates the proportion of word categories between AD and healthy control people. Our review of the literature indicates that this report is the first quantitative study investigating the word categories associated with AD in Japanese. The statistics is presented in Table 1.

Ummm, ummm, what were those?
... Oh no.
Well, they were all food, weren't they?
A plant?
A plant... a plant... was *it* a cherry blossom?

 (Omitted)

What were the *other* 2? There was a pencil, a watch and what else? ... I cannot remember *it*. Ah, well, a key. And there was, there was a pencil here. And that one. What was the 3rd one here? ...... There was a key. ...... There was a key and I cannot remember the next one. ... I don't know. What was *it*? Tell me.

Figure 1: Speech of a patient with AD in Mini-Mental State Examination (MMSE score) (Translated into Japanese). *Italic* indicates personal impersonal pronouns.

Maximum	1569
Minimum	242
Median	688
Average	788

Table1: Word Statistics in Corpus.

Contributions of this study can be summarized as shown below.
● 	A LIWC analysis is conducted for narratives uttered by people suspected of having AD.
● 	This study also examines a proposed method for LIWC translation.

## 2 	Related Work

Recent studies of early detection methods such as blood testing and detailed memory testing have revealed vast improvements in detection capabilities (Mapstone et al. 2014). However, most of these methods are physically or mentally invasive, which has led to anticipation of less-invasive or even non-invasive detection methods. Dementia symptoms include degenerative cognitive decline, as well as behavioral and functional disorders. The disease also results in the deterioration of various executive functions, reasoning, and language abilities. Among these, language deficits have been shown to be more apparent from the early stages of dementia (Snowdon et al. 1996). These deficits include naming disorders, auditory and written comprehension impairment, fluent but empty speech, and semantic paraphasia. However, repetition capabilities and articulation are often preserved (Appell et al. 1982). Reportedly, the impairment of language abilities in dementia patients is often inconsistent because semantic and pragmatic language abilities are likely to become more impaired, whereas syntax and phonology demonstrate better preservation (Schwartz et al. 1979). Semantic errors reportedly are

---

[1] http://www.businessinsider.com/changes-in-president-reagans-speech-early-sign-of-alzheimers-2015-4

the most common and distinct language deficit because dementia patients tend to substitute target names with superordinate category names or demonstrate circumlocutory speech with impaired naming (Emery 2000). Other reports have also described unrelated errors (Moreaud et al. 2001), phonological errors (Croot et al. 2000), and visual errors (Croot et al. 2000). However, these are often dependent on the type of picture confrontation naming task, the severity or stage of the disease, or other unique patient-level circumstances (Geda 2012). MCI, part of which constitutes a pre-stage of dementia, might indicate a boundary between aging-related non-dementia reduction in cognition and dementia on the spectrum of cognitive function.

Using the above characteristics, various dementia screening methods have been proposed to date. Table 2 shows the summary of previous screening studies. Well-known studies were those conducted by Roark et al. (Roark et al. 2007; Roark et al. 2011), which analyzed the lexical features and syntactic feature from transcripts of spoken narrative such as neuropsychological approaches (Moriyama et al. 2015) and automatic speech analysis approaches (König et al. 2015). Some of them used automatic speech recognition (Tóth et al. 2015). Aramaki et al. specifically examined vocabulary size in speech transcription (Aramaki et al. 2016). Tanaka et al. proposed a novel approach using computer avatars (Tanaka et al. 2016). In addition Orimaye et al. (Orimaye et al. 2014) used machine learning algorithms to build diagnostic models using syntactic and lexical features and Jarrold et al. used LIWC for aided diagnosis of Dementia (Jarrold et al. 2014).

Author	Method	Disease	Sample size	Year
Aramaki et al. (Aramaki et al. 2016)	Analysis of vocabulary size in speech	MCI, AD	22	2016
Tanaka et al. (Tanaka et al. 2016)	Spoken dialog with computer avatars	MCI	18	2016
König et al. (König et al. 2015)	Automatic speech analyse	MCI, AD	64	2015
Tóth et al. (Tóth et al. 2015)	Acoustic indicator	MCI	51	2015
Moriyama et al. (Moriyama et al. 2015)	Neuropsychological battery	AD	299	2015
Orimaye et al. (Orimaye et al. 2014)	Machine learning algorithms	AD	556	2014
Jarrold et al. (Jarrold et al. 2014)	Analysis of spontaneous speech	AD	48	2014
Roark et al. (Roark et al. 2011)	Transcript with audio	MCI	74	2011
Roark et al. (Roark et al. 2007)	Lexical features and syntactic features	MCI	55	2007

Table 2: Earlier studies.

3   Materials

We have collected narratives of hospital patients to build the corpus.

3.1   Research field

Criteria used for the experiment are the following.
[Inclusion criterion]
- **AD group (AD)**: Patients with Alzheimer's disease between light MCI and middle class MCI (MMSE below 21 points).
- **Healthy control group (HC)**: Patients without AD. Healthy control people group members are age-matched with AD group members (MMSE over 22 points)[2].

[Exclusion criteria]
- Patients who have some other brain-related diseases
- Non-native Japanese speakers

We recorded conversations between a patient and a medical staff member using an IC recorder. Then, we transcribed the conversations manually. Table 3 presents characteristics of the patients.

---

[2] Healthy control people might actually have diseases (other than AD).

Age	Sex	MMSE		Age	Sex	MMSE
72	Woman	4		77	Man	22
71	Woman	14		81	Woman	22
90	Woman	17		72	Woman	22
80	Man	18		87	Man	25
73	Woman	19		53	Man	25
78	Woman	19		87	Woman	26
81	Woman	20		82	Woman	26
73	Man	21		79	Man	28
77	Woman	21		71	Woman	30
(a) AD				(b) HD		

Table 3: Participant attributes.

3.2 Ethics Statement

The experiment is explained to patients (or their family). If they cannot understand the explanation, then we exclude them. We do not reward the patients. The use of these data for research purposes was approved by the ethics committee of Kyoto University (approval NO. E2525).

3.3 MMSE – based Patient Classification

The goal of this study is to detect important features that can classify AD and others by analyzing their spoken narratives. Thus, we determine a person is AD or not (MCI and normal) based on Mini Mental State Examination (MMSE). The MMSE is a simple inspection method for a subject suspected as AD. In this test, a patient is asked 11 questions; their answers are judged by the score (max 30; min 0). The MMSE scores between 30 and 27 points are normal; those between 26 and 22 points might be MCI; and those below 21 points might be AD.

4 Language Resource LIWC

4.1 What is LIWC

We use Linguistic Inquiry and Word Count (LIWC) as a language resource for classifying words into corresponding categories. LIWC has been developed by researchers who are interested in social, clinical, health, and cognitive psychology. We can classify people's social and psychological states using LIWC. However because LIWC is only for English, it is difficult to apply to Japanese texts. Our review of literature indicates that no resource for Japanese is comparable with LIWC.

Therefore, we make Japanese LIWC by translating English LIWC. We arrange categories for Japanese LIWC by considering a gap depending on the language differences. Table 4 shows 64 categories in English LIWC. Then we extract 22 categories related to diseases by the judgment of the authors, as shown in Table 5. We remove categories that are not related to our target disease (e.g. <Body>). We also remove categories that are not translatable to Japanese. (e.g. <Article>).

<Funct>	<Ipron>	<Sad>	<Incl>	<Adverbs>	<Family>	<Body>	<Work>
<Pronoun>	<Article>	<CogMech>	<Excl>	<Prep>	<Friends>	<Health>	<Achiev>
<Ppron>	<Verbs>	<Insight>	<Percept>	<Conj>	<Humans>	<Sexual>	<Leisure>
<I>	<AuxVb>	<Cause>	<See>	<Negate>	<Affect>	<Ingest>	<Home>
<We>	<Past>	<Discrep>	<Hear>	<Quant>	<Posemo>	<Relativ>	<Money>
<You>	<Present>	<Tentat>	<Feel>	<Numbers>	<Negemo>	<Motion>	<Relig>
<SheHe>	<Future>	<Certain>	<Bio>	<Swear>	<Anx>	<Space>	<Death>
<They>	<Filler>	<Inhib>	<Nonflu>	<Social>	<Anger>	<Time>	<Assent>

Table 4: English LIWC categories. (64 categories)

<Time>	<Posemo>	<Ipron>	<Sad>	<Family>	<Negemo>	<Present>	<Humans>
<Future>	<Space>	<Anger>	<Negate>	<SheHe>	<I>	<Friends>	
<Social>	<Past>	<Verbs>	<We>	<Anx>	<They>	<You>	

Table 5: Disease - related categories. (22 categories)

4.2 LIWC Translation Procedure

We translate LIWC into Japanese to produce Japanese LIWC as shown below.

- **Step 1**: All words in English LIWC were translated using EDICT (an electric dictionary produced by EDP and JimBreen[3]).
- **Step 2:** One worker searched mistakenly translated words by sight and deleted them. As a result, 5,534 words out of 6,211 words remained.
- **Step 3**: If a duplicated pairs of a Japanese word and its category are found, then we put them together such that 4,769 words out of 5,534 words remained.
- **Step 4**: When conducting morphological analysis for Japanese, we ignore words in the category `<Past>`. Then the words of verbs belonging to the category `<Present>` is changed to `<Verbs>`. We remove three categories `<We>`, `<SheHe>`, and `<They>` and words belonging to these because it was determined that these categories have no correlation with disease.
  Therefore, the number of categories are reduced from 22 to 19.
- **Step 5**: Words in multiple categories are assigned to the most appropriate category by one worker. In Japanese, it is difficult to distinguish between words related to `<Time>` and those related to `<Space>`. Therefore, we define a new category called `<TimeSpace>`. The number of categories becomes 20. We apply these steps to 2,700 words.

5　Experiments

5.1　Procedure

We analyze the corpus as explained below.

- **Step 1**: Texts are analyzed morphologically and stemmed using a Japanese morphological analyzer (Kurohashi and Nagao 2003).
- **Step 2**: The results are consulted by Japanese LIWC. We then count the LIWC word in the corpus.
- **Step 3**: We investigate the ratio of LIWC word frequency for each category.

5.2　Results

The results of *t*-test are presented in Table 6. In order to the examine the difference between speech of AD group and HC group in a statistical manner. Note that we investigated the difference of the average values in AD and HC group. As shown in Table 6, no significant difference was found between AD and HC in any categories, except for four: `<Social>`, impersonal pronoun `<Ipron>`, anxiety `<Anx>`, `<Verbs>`, and `<Present>`. As for `<Anx>`, HC's value is 0. Figure 2 presents results of the category frequency of AD and HC.

Category	AD (avg.)	HC (avg.)	*p-value*	Difference
`<Ipron>`	**0.0385**	**0.0268**	**0.0187**	0.0117
`<Anx>`	0.0008	0	0.0192	0.0008
`<Verbs>`	**0.0524**	**0.043**	**0.0219**	0.0094
`<Present>`	**0.0171**	**0.0103**	**0.0226**	0.0068
`<Social>`	**0.0063**	**0.0116**	**0.0229**	-0.0053
`<I>`	0.004	0.0019	0.0591	0.0021
`<Space>`	0.017	0.0231	0.0893	-0.0061
`<Posemo>`	0.006	0.0076	0.1245	-0.0016
`<Time>`	0.0364	0.0418	0.1433	-0.0054
`<Sad>`	0	0.0002	0.1733	-0.0002
`<You>`	0.0003	0.0002	0.2687	0.0001
`<Family>`	0.0015	0.0021	0.3135	-0.0006
`<Negate>`	0.0397	0.0464	0.3264	-0.0067
`<Negemo>`	0.0006	0.0009	0.3294	-0.0003
`<Anger>`	0.0004	0.0006	0.3392	-0.0002
`<Humans>`	0.0068	0.0077	0.3432	-0.0009
`<Friends>`	0.0008	0.0006	0.4019	0.0002
`<Past>`	0.0003	0.0003	0.4909	0
`<Future>`	0	0	-	0
`<TimeSpace>`	0	0	-	0

Table 6: Values that has significant differences between AD and HC (*p*-value < 0.05) are **under lined**.
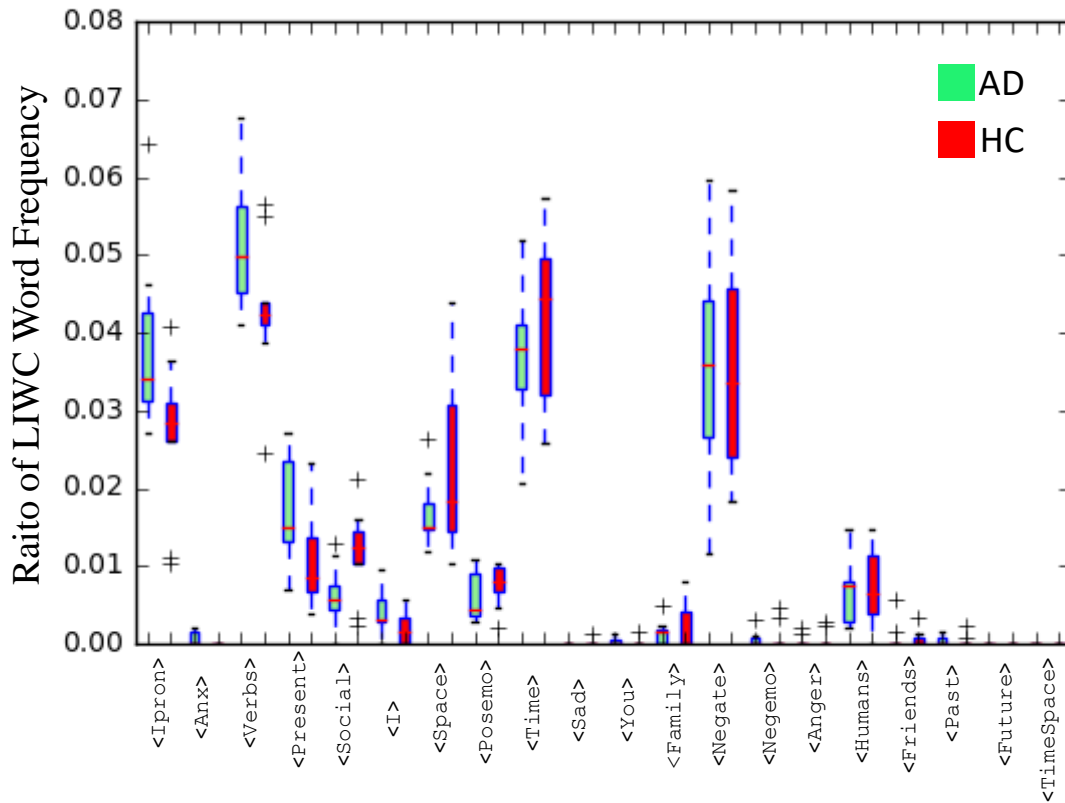
Figure 2: Category frequency of AD (green) and HC (red). Significant differences were found for
`<Ipron>`, `<Anx>`, `<Verbs>`, `<Present>`, and `<Social>`.

6 Discussion

First, we discuss the findings based on quantitative evidence obtained from a previous study (Sec. 6.1).
Then we examine the results by using machine learning (Sec. 6.2).

6.1 Findings: Quantitative Evidence of Previous Study

We discuss categories for which significant differences between AD and HC are observed. The values
of `<Social>` in AD group were significantly lower than those in HC group. Generally, it is said that
participating in social activities is effective to prevent AD progression. In other words, a person with
little social contact tends to develop AD. Consequently, this result corresponds with AD features. The
values of `<Ipron>` in AD group were significantly higher than those in HC group. AD patients be-
come forgetful. Therefore, they use many impersonal pronouns (Almor et al. 1999). Viewed from a
grammatical perspective, ellipses of a subject or objects of a verb are not allowed. They often appear
as a pronoun in English, but the ellipsis of them is allowed in the Japanese language. Considering this
feature, it is possible that the use of impersonal pronouns becomes more frequent in the condition of
AD, particularly for Japanese speakers. Similarly, it corresponds with general AD features. The values
of `<Verbs>` and `<Present>` in AD group were also significantly larger than those in HC group.
However, it is difficult to understand why these results were obtained. Therefore, in future work, it
will be necessary to investigate the words in these categories in detail.

   Consequently, some observed results supported the previous findings on AD. Although most of the
previous studies have been based on subjective observations, our findings provide quantitative evi-
dence for their claims, demonstrating the effectiveness of our approach.

6.2 Decision Tree

In order to the most important clue to classify patients into AD and HC, a decision tree is constructed as shown in Figure 3. It has feature values representing probabilities to be classified into AD or HC in categories.

```
                          <Ipron> <= 0.0298
                              n = 18
                            AD: 9 HC: 9
              Yes                              No

        <Posemo> <= 0.0046              <Space> <= 0.0285
             n = 7                           n = 11
          AD: 1 HC: 6                     AD: 8 HC: 3

    n = 1          n = 6        <Posemo> <= 0.0025        n = 2
  AD: 1 HC: 0    AD: 0 HC: 6         n = 9            AD: 0 HC: 2
                                  AD: 8 HC: 1

                            n = 1          n = 8
                          AD: 0 HC: 1    AD: 8 HC: 0
```
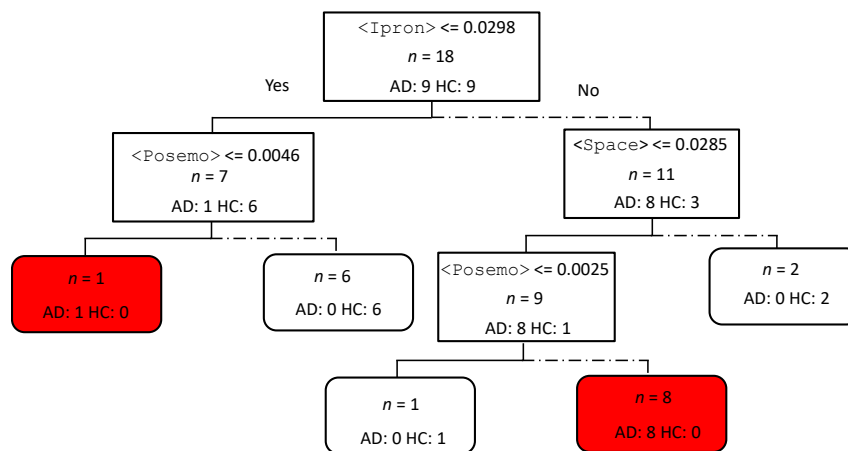
Figure 3: Decision tree results for AD screening.

Figure 3 shows that there are two cases for a person to be diagnosed as AD. The first case is that of using words in <Ipron> below 0.0298 and the value using words in <Posemo> below 0.0046. The probability is 100% to be classified into AD. The other case is that the percentage of impersonal pronouns <Ipron> is higher than 0.0298, the percentage of Space <Space> is less than 0.0285 and the percentage of Positive emotion <Posemo> is higher than 0.0025. The probability is also 100% for classification into AD. Results demonstrate that the values of appearance of words of <Ipron>, <Space> and <Posemo> in conversation are important for AD screening.

7 Conclusion

This study investigated features of the words that AD patients used in their utterances. 18 examinees of 53–90 years old (mean: 76.89) were recruited and divided into two groups based on their MMSE scores. Linguistic Inquiry and Word Count (LIWC) classified words were used to categorize the words that the examinees used. Then their frequency was ascertained. This report is the first of a quantitative study that investigated the word categories of AD. Significant differences were found for the AD group in the usage of several LIWC categories, including impersonal pronouns, which suggests that this simple method can be used for dementia screening.

References

Appell, J., A. Kertesz and M. Fisman (1982). "A study of language functioning in Alzheimer patients." *Brain and Language* 17(1): 73-91.

Aramaki, E., S. Shikata, M. Miyabe and A. Kinoshita (2016). "Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment." *PLoS One* 11(5): e0155195.

Croot, K., J. R. Hodges, J. Xuereb and K. Patterson (2000). "Phonological and articulatory impairment in Alzheimer's disease: a case series." *Brain and Language* 75(2): 277-309.

Emery, V. O. B. (2000). "Language impairment in dementia of the Alzheimer type: a hierarchical decline?" *The International Journal of Psychiatry in Medicine* 30(2): 145-164.

Geda, Y. E. (2012). "Mild cognitive impairment in older adults." *Current Psychiatry Reports* 14(4): 320-327.

König, A., A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten and P. H. Robert (2015). "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease." *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 1(1): 112-124.

Kurohashi, S. and M. Nagao (2003). Improvements of Japanese Morphological Analyzer JUMAN. Treebanks, *Springer*.

Mapstone, M., A. K. Cheema, M. S. Fiandaca, X. Zhong, T. R. Mhyre, L. H. MacArthur, W. J. Hall, S. G. Fisher, D. R. Peterson and J. M. Haley (2014). "Plasma phospholipids identify antecedent memory impairment in older adults." *Nature Medicine* 20(4): 415-418.

Moreaud, O., D. David, A. Charnallet and J. Pellat (2001). "Are semantic errors actually semantic?: Evidence from Alzheimer's disease." *Brain and Language* 77(2): 176-186.

Moriyama, Y., A. Yoshino, K. Yamanaka, M. Kato, T. Muramatsu and M. Mimura (2015). "The Japanese version of the Rapid Dementia Screening Test is effective compared to the clock-drawing test for detecting patients with mild Alzheimer's disease." *Psychogeriatrics*.

Orimaye, S. O., J. S.-M. Wong and K. J. Golden (2014). Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances. *Proceedings of the First Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*.

Roark, B., M. Mitchell and K. Hollingshead (2007). Syntactic complexity measures for detecting mild cognitive impairment. *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, Association for Computational Linguistics.

Roark, B., M. Mitchell, J. P. Hosom, K. Hollingshead and J. Kaye (2011). "Spoken Language Derived Measures for Detecting Mild Cognitive Impairment." *IEEE Trans Audio Speech Lang Process* 19(7): 2081-2090.

Schwartz, M. F., O. S. Marin and E. M. Saffran (1979). "Dissociations of language function in dementia: A case study." *Brain and Language* 7(3): 277-306.

Snowdon, D. A., S. J. Kemper, J. A. Mortimer, L. H. Greiner, D. R. Wekstein and W. R. Markesbery (1996). "Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study." *JAMA* 275(7): 528-532.

Tanaka, H., H. Adachi, N. Ukita, T. Kudo and N. Satoshi (2016). Automatic Detection of Very Early Stage of Dementia through Spoken Dialog with Computer Avatars. *IEEE Engineering in Medicine and Biology Society*.

Tóth, L., G. Gosztolya, V. Vincze, I. Hoffmann and G. Szatlóczki (2015). Automatic detection of mild cognitive impairment from spontaneous speech using ASR, *ISCA*.

Wortmann, M. (2015). World Alzheimer report 2014: Dementia and risk reduction. Alzheimer's & Dementia: *The Journal of the Alzheimer's Association* 11(7): P837.

Jarrold, W., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M. L., & Ogar, J. (2014). Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology* : pages 27-36.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and language*, *67*(3), 202-227.