


## RESEARCH ARTICLE

# Reversal of the concreteness effect can be detected in the natural speech of older adults with amnestic, but not non-amnestic, mild cognitive impairment

Luwen Cao<sup>1</sup> | Kunmei Han<sup>1</sup> | Li Lin<sup>1,2</sup> | Jiawen Hing<sup>1</sup> | Vincent Ooi<sup>1</sup> |  
 Nick Huang<sup>1</sup> | Junhong Yu<sup>3</sup> | Ted Kheng Siang Ng<sup>4,5</sup> | Lei Feng<sup>4,6,7</sup> |  
 Rathi Mahendran<sup>4</sup> | Ee Heok Kua<sup>4</sup> | Zhiming Bao<sup>1,8</sup> 

<sup>1</sup>Department of English, Linguistics and Theatre Studies, National University of Singapore, Singapore, Singapore

<sup>2</sup>School of Foreign Studies, East China University of Political Science and Law, Shanghai, China

<sup>3</sup>Cognitive and Brain Health Laboratory, School of Social Sciences, Nanyang Technological University, Singapore, Singapore

<sup>4</sup>Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

<sup>5</sup>Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, USA

<sup>6</sup>Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

<sup>7</sup>Centre for Healthy Longevity, Clinic L, Alexandra Hospital, Singapore, Singapore

<sup>8</sup>Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China

## Correspondence

Zhiming Bao, Department of English,  
 Linguistics and Theatre Studies, National  
 University of Singapore, 7 Arts Link, Singapore.  
 Email: [ellbaozm@nus.edu.sg](mailto:ellbaozm@nus.edu.sg)

## Funding information

Humanities and Social Sciences Research  
 Fund, National University of Singapore; Social  
 Science Research Thematic Grant, Ministry of  
 Education, Singapore; DSO National  
 Laboratories, Singapore; Kwan Im Thong Hood  
 Cho Temple, Singapore; Lee Im Tah Holdings,  
 Singapore

## Abstract

**INTRODUCTION:** Patients with Alzheimer's disease present with difficulty in lexical retrieval and reversal of the concreteness effect in nouns. Little is known about the phenomena before the onset of symptoms. We anticipate early linguistic signs in the speech of people who suffer from amnestic mild cognitive impairment (MCI). Here, we report the results of a corpus-linguistic approach to the early detection of cognitive impairment.

**METHODS:** One hundred forty-eight English-speaking Singaporeans provided natural speech data, on topics of their choice; 74 were diagnosed with single-domain MCI (38 amnestic, 36 non-amnestic), 74 cognitively healthy. The recordings yield 267,310 words, which are tagged for parts of speech. We calculate the per-minute word counts and concreteness scores of all tagged words, nouns, and verbs in the dataset.

**RESULTS:** Compared to controls, subjects with amnestic MCI produce fewer but more abstract nouns. Verbs are not affected.

**DISCUSSION:** Slower retrieval of nouns and the reversal of the concreteness effect in nouns are manifested in natural speech and can be detected early through corpus-based analysis.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

**KEYWORDS**

Alzheimer's disease, imageability, mild cognitive impairment, natural speech, reversal of the concreteness effect

**Highlights**

- Reversal of the concreteness effect is manifested in patients with Alzheimer's disease (AD) and semantic dementia.
- The paper reports a corpus-based analysis of natural speech by people with amnesic and non-amnesic mild cognitive impairment (MCI) and cognitively healthy controls.
- People with amnesic MCI produce fewer and more abstract nouns than people with non-amnesic MCI and healthy controls. Verbs appear to be unaffected.
- The imageability problem can be detected in natural everyday speech by people with amnesic MCI, which carries a higher risk of conversion to AD.

## 1 | BACKGROUND

In recent years, there has been a growing body of research on the distinct roles that nouns and verbs play in cognitive impairment. It has been well documented that patients diagnosed with semantic dementia (semantic variant of frontotemporal primary progressive aphasia) or Alzheimer's disease (AD) present with effortful lexical retrieval and reversal of the concreteness effect, producing nouns, and to a lesser extent, verbs, which are more abstract.<sup>1-9</sup> The matter is, of course, controversial. There are studies that report contrasting results, namely, reverse concreteness effects are not common in semantic dementia,<sup>10,11</sup> and imageability plays no or little role in AD.<sup>12,13</sup> Most studies collect targeted language data from word-based fluency tests on semantic categories (*cat, dog*) or letters (*cat, cake*), or from connected speech obtained through structured interviews, picture narrations (e.g., *Cookie Theft*) or fairy tale recalls (e.g., *Cinderella*).<sup>14,15</sup> The datasets of these studies are relatively small, even those constructed from connected speech, and the speech data are constrained by visual or reading stimuli and by research designs. To our knowledge, there has been no or little study of per-minute counts and concreteness of nouns and verbs in patients with mild cognitive impairment (MCI) based on unstructured and easy-to-collect speech data. In the present study, we take a corpus and formal linguistic approach to search for linguistic markers of early cognitive impairment in natural, unconstrained speech by people with mild impairment in the cognitive domain of memory, by people with mild impairment in cognitive domains other than memory, and by people who are cognitively healthy.

## 2 | METHODS

### 2.1 | Neuropsychological assessment

Our language dataset comes from participants in the Community Health and Intergenerational (CHI) study, a cohort study of aging and

mental health among normal, community-dwelling Singaporeans  $\geq 60$  years of age. A total of 993 Singaporeans participated in the CHI study. The aims and methods of the CHI project have been described in detail elsewhere.<sup>16</sup> Assessments include physical health, socioeconomic conditions, cognitive functioning, and unconstrained speech. The neuropsychological battery of tests used to diagnose normal aging and MCI in the study has been validated in the Singaporean population.<sup>17</sup> The tests evaluate amnesic and non-amnesic cognitive domains of attention, learning, memory, speed, and executive function, which are necessary for the diagnosis of neurocognitive disorders. The battery of five tests used in the cognitive assessment is (1) Rey Auditory Verbal Learning Test (RAVLT), to evaluate declarative verbal learning and memory; (2) immediate, delayed, and recognition memory tests and forward and backward digit span tasks, to assess attention and verbal working memory; (3) Color Trails Tests 1 and 2, to assess sustained attention and reasoning<sup>18</sup>; (4) block design, to measure visual-spatial and organizational processing abilities and non-verbal problem-solving skills; and (5) semantic verbal fluency (animal) test, to tap lexical knowledge and semantic memory organization.<sup>19</sup> We used Petersen's criteria<sup>20</sup> for the diagnosis of amnesic and non-amnesic MCI, that is, absence of functional decline and  $\geq 1$  neuropsychological test score ( $-1.5$  standard deviation [SD] below the norm). Specifically, an individual was diagnosed with MCI under three conditions: first, at least one impaired test score in the above-mentioned neuropsychological tests, defined as 1.5 SD below their age-appropriate norms; second, largely preserved functional independence as defined by having a Clinical Dementia Rating global score  $< 1$ ; third, presence of subjective cognitive complaints as corroborated by a reliable informant. For MCI subtypes, participants were diagnosed with amnesic MCI if they had at least one impaired test score (i.e., 1.5 SD below age-appropriate norms) from the memory domain (e.g., immediate recall, delayed recall, recognition). Non-amnesic MCI was diagnosed if at least one test score from the non-memory domains was impaired, in the absence of any impaired memory test scores. The diagnosis of MCI for each subject was arrived at through a consensus review of the test results by two

**TABLE 1** The number of subjects diagnosed with impairment of a single non-amnestic domain, as determined by the neurocognitive tests.

Test	N
Block design	7
Color trails 1	1
Color trails 1 & 2	11
Digit span	2
Semantic fluency	15

Note: Total number of subjects: 36.

psychiatrists and a neuropsychologist. Subtyping for amnestic and non-amnestic MCI has relevance for clinical management and prognosis. Amnestic MCI is associated with higher risk of conversion to AD,<sup>21</sup> whereas non-amnestic MCI is associated with higher risk of conversion to other types of dementia such as Lewy body dementia.<sup>22</sup>

## 2.2 | Selection criteria

Of the 993 participants in the CHI study, 475 subjects in their 60s and 70s provided speech samples in English, including 95 who were diagnosed with MCI. Of the 95 MCI subjects, 53 suffer from amnestic MCI (38 single-domain, 15 multiple-domain) and 42 from non-amnestic MCI (36 single-domain, 6 multiple-domain). To minimize the effect of other cognitive domains on language production, we selected all 38 subjects with single-domain amnestic MCI (aMCI) and all 36 subjects with single-domain non-amnestic MCI (naMCI). We also selected 74 cognitively healthy participants with similar age, sex, education, and language profiles. The selection was random, and no other criterion was used. The non-amnestic domains examined correspond to the neurocognitive tests used for assessing the non-memory domains, such as working memory (digit span), processing speed (color trails 1 and 2), divided attention (color trails 2), visual-spatial processing (block design), and verbal fluency (semantic fluency). The number of subjects diagnosed with mild impairment in individual non-amnestic domains is shown in Table 1.

Because the number of subjects is small, we grouped all 36 naMCI subjects as a single group. Future research could build on our data to investigate the linguistic manifestation of individual non-amnestic domains. The neurocognitive profile of these MCI subjects compared to controls is illustrated in Figure 1.

All the neurocognitive tests, except for the RAVLT (learning, delayed recall, and recognition), belong to the non-amnestic cognitive domains.

## 2.3 | Natural speech data

Most Singaporeans are bilingual, speaking English and one or more of the heritage languages of Chinese, Malay, or Tamil. For most, English is the dominant home language.<sup>23</sup> Participants were instructed to talk about any topic for up to 20 minutes in a language they felt most

## RESEARCH IN CONTEXT

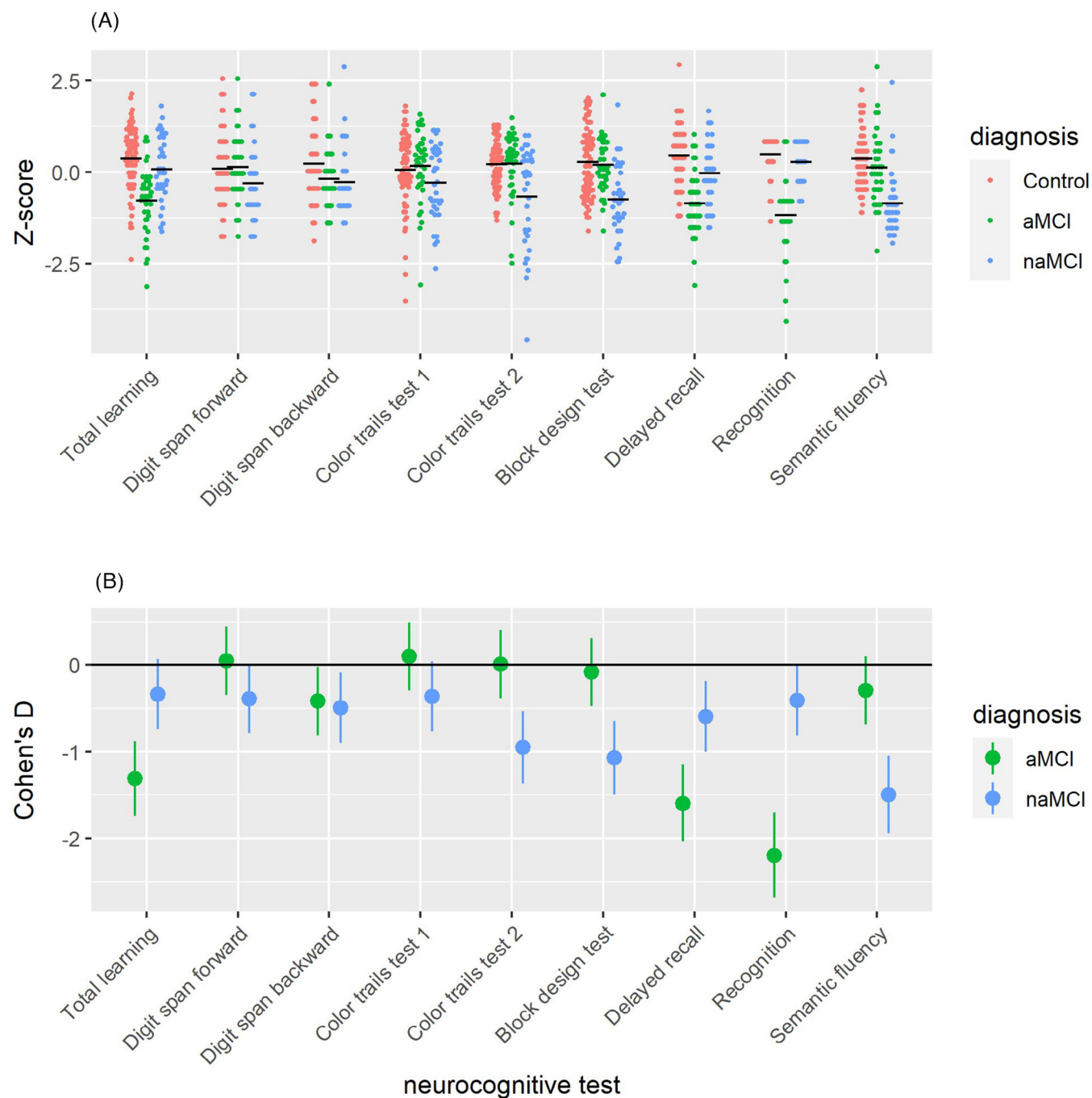
- 1. Systematic review:** Reversal of the concreteness effect is associated with Alzheimer's disease. Current work on the linguistic manifestation of cognitive impairment relies on language data collected from controlled settings, such as picture description and fairy tale recall. The authors collected natural speech data from 148 English-speaking Singaporeans as part of a cohort study of aging, 74 diagnosed with single-domain mild cognitive impairment (MCI; 38 amnestic, 36 non-amnestic) and 74 cognitively healthy.
- 2. Interpretation:** We tagged the words for parts of speech and calculated the per-minute counts and the concreteness scores of nouns and verbs. Compared to controls, people with amnestic MCI produce fewer and more abstract nouns. Verbs are spared. Natural speech data can be used for early detection of amnestic MCI.
- 3. Future directions:** If confirmed in future studies with larger population bases, we can investigate the linguistic manifestation of individual non-amnestic domains, and devise early linguistic intervention strategies to address the imageability problems afflicting people with amnestic MCI.

comfortable with, with minimal involvement from interviewers. The speeches were recorded with simple digital voice recorders in an ordinary office setting. Topics varied freely and widely, ranging from work and retirement to family life and public affairs.

The recordings were transcribed verbatim by Singaporean students at the National University of Singapore who are familiar with the local languages. The transcribers were not involved in the initial recording sessions. We used the Stanford PoS tagger to tag the transcribed words for parts of speech, based on the Penn Treebank tagset.<sup>24,25</sup> The tagged words were manually vetted by another group of students trained in formal linguistics and in the descriptive grammar of English. The raw recordings contained interviewer–subject interactions. For each recording, we removed the words uttered by the interviewer as prompts, encouragements, and explanations, and adjusted the time of the recording accordingly. We also removed the time for pauses and hesitations during the interactions between the interviewer and the subject, but kept the subject's own pauses, repetitions, and false starts during their continuous speech. After the adjustment, the remaining time of a recording is the net talk time of the subject.

## 2.4 | Linguistic variables

We calculated noun and verb densities as a percentage of all words, and type-token ratio (TTR) as a moving average with a 20-word window.<sup>26</sup>



**FIGURE 1** The neurocognitive profile of MCI subjects in comparison to controls as illustrated in the distribution of a) standardized scores and b) Cohen's D statistics characterizing the effect sizes of the between-group difference. To facilitate interpretation, the scores are z-standardized across the full study sample for each of the neurocognitive tests. Additionally, the scores for Color trails test 1 and 2 were multiplied by  $-1$ , such lower Z-scores and more negative Cohen's Ds represent greater impairment in these tests, in line with the other tests.

Lexical densities and TTR are common measures of lexical diversity and richness in corpus linguistics.

It is well documented that people with aMCI experience difficulty in lexical retrieval, especially of nouns which encode semantic knowledge.<sup>27</sup> Because the lexical measures of noun and verb densities do not differentiate among aMCI, naMCI, and control, we investigated the usage patterns of nouns and verbs per minute of talk time in the

speech of the three groups. For each subject's speech sample, we calculated the per-minute counts of nouns and verbs by dividing the total number of noun or verb tokens by the total number of minutes. We calculated the concreteness scores of the speech sample of each subject based on the concreteness ratings of 40,000 English words compiled by Brysbaert et al.,<sup>28</sup> with one being the most abstract and five being the most concrete. For an inflected word which does not have an

**TABLE 2** The means (standard deviations) in age, sex, education, and language of the subjects who volunteered to provide natural speech data.

	Control	aMCI	naMCI	F(2,145)	P
N	74	38	36		
Age	66.3 (4.8)	65.1 (4.1)	66.4 (5.7)	0.893	0.412
Sex (M/F)	37/37	24/14	13/23	$\chi^2 = 5.41$	0.067
Education	14.8 (3.2)	13.8 (3.1)	13.9 (4.2)	1.334	0.267
Language	2.3 (0.5)	2.2 (0.6)	2.3 (0.5)	1.276	0.282

Note: The means of age and education are in years, and the mean of language is in the number of languages spoken.  
Abbreviations: aMCI, amnesic mild cognitive impairment; naMCI, non-amnesic mild cognitive impairment.

independent score, we used the score of the lemma (*pass* for *passed*, *colleague* for *colleagues*). For each word or lemma, we obtained the score from the 40,000-word database, and multiplied it by the number of tokens of the word, to arrive at the word's token score. The sum of all word token scores was divided by the total number of word tokens to yield the concreteness score of the speech sample of each subject. Approximately 3.4% of common noun lemmas and 0.6% of verb lemmas in our speech samples do not appear in the database and are not rated. Also not rated are proper nouns (*Singapore*, *Singaporean*), foreign-origin words (*tau* "soybean") and sentence-final particles unique to Singapore English (*lah*). These words were not included in the calculation. We also excluded numerals and common nouns used as proper nouns (*Elm Street*).

## 2.5 | Statistical analysis

All the variables used in the study, except sex, are continuous. For statistical analysis, we use the one-way analysis of variance for the continuous variables, with post hoc Tukey contrast, and the chi-square test for the categorical variable. The tests are performed with SPSS v.28.

## 3 | RESULTS

### 3.1 | Basic demographics of subjects

The basic demographic information of the subjects who contributed speech samples to the dataset is shown in Table 2.

For most of the 148 subjects, Chinese is their second language; six are English–Malay or English–Tamil bilingual, and five monolingual. There is no significant difference between the subjects with the two subtypes of MCI and cognitively healthy controls in age, education, and language.

### 3.2 | Lexical profiles

Table 3 displays the basic lexical statistics of the dataset.

People with aMCI and naMCI talk less, and produce fewer words, than cognitively healthy controls. The differences in talk time and in the

three-word count measures, both type and token, are statistically significant at  $P < 0.05$ . The differences between the two MCI groups are not significant, as revealed by the post hoc Tukey contrasts in Table 3. As can be seen from Table 3, there is no difference in the three lexical measures, noun density, verb density, and TTR, in the language samples of the subjects in the aMCI, naMCI, and control groups. In fact, the TTRs of the three groups are nearly identical with the TTR, 0.85 (0.03), of the control group, consisting of 37 healthy Americans with similar age and education profiles, in the studies of the lexical features of frontotemporal degeneration reported by Cho et al.<sup>8,9</sup>

### 3.3 | Lexical retrieval and concreteness

The per-minute word counts and concreteness scores of nouns and verbs are shown in Table 4.

As we can see in Table 4, the all-word counts of the three groups exhibit significant variation ( $P = 0.004$ ), and the post hoc Tukey test shows that only the naMCI group is significantly slower than the control group (141.9 vs. 126.2,  $P = 0.006$ ). The concreteness scores of the three groups do not show significant variation ( $P = 0.145$ ). When we examine nouns and verbs separately, we will be able to see important differences between the two major lexical categories obscured by the all-word measures. In the case of word count per minute, nouns and verbs behave differently between the two subtypes of MCI. Compared to control, the per-minute word counts of nouns and verbs were lower in both subtypes, but only the lower noun count in aMCI is significant (24.3 vs. 22.1,  $P = 0.038$ ). The concreteness scores of nouns and verbs exhibit similar trends. Compared to control, nouns are more abstract in both subtypes of MCI; again, only the lower noun score in aMCI is significant (3.71 vs. 3.57,  $P = 0.027$ ). Verbs show no significant differences. Taken together, on the two measures of per-minute word count and concreteness score, our data suggest that amnesic MCI affects nouns and spares verbs, and non-amnesic MCI does not seem to affect per-minute word count nor concreteness.

## 4 | DISCUSSION

The results, and the dissociation between nouns and verbs, are consistent with the findings reported in the linguistic and neuropsychological



**TABLE 3** Means (standard deviations) of talk time (minutes) and type and token counts (words) of all tagged words, nouns, verbs, and the three lexical measures of noun density, verb density, and type-token ratio (TTR) in the dataset.

	Control	aMCI	naMCI	F(2,145)	P	Post hoc Tukey contrasts
N (M/F)	37/37	24/14	13/23			
Talk time	14.4 (3.9)	12.6 (6.0)	11.4 (4.9)	5.271	0.006	Control > aMCI ( $P = 0.143$ ), Control > naMCI ( $P = 0.006$ ), aMCI > naMCI ( $P = 0.520$ )
Words, type	523.7 (128.1)	448.4 (166.5)	422.1 (146.2)	7.321	0.001	Control > aMCI ( $P = 0.025$ ), Control > naMCI ( $P = 0.002$ ), aMCI > naMCI ( $P = 0.709$ )
Words, token	2038.7 (664.5)	1661.3 (897.1)	1481.0 (811.4)	7.333	0.001	Control > aMCI ( $P = 0.039$ ), Control > naMCI ( $P = 0.001$ ), aMCI > naMCI ( $P = 0.571$ )
Nouns, type	172.6 (60.7)	137.4 (64.1)	132.0 (54.2)	7.416	0.001	Control > aMCI ( $P = 0.011$ ), Control > naMCI ( $p = 0.003$ ), aMCI > naMCI ( $p = 0.923$ )
Nouns, token	351.2 (129.2)	278.5 (141.8)	258.4 (133.2)	7.268	0.001	Control > aMCI ( $p = 0.019$ ), Control > naMCI ( $P = 0.002$ ), aMCI > naMCI ( $P = 0.795$ )
Verbs, type	147.9 (38.9)	129.4 (52.7)	119.2 (48.1)	5.476	0.005	Control > aMCI ( $P = 0.103$ ), Control > naMCI ( $P = 0.006$ ), aMCI > naMCI ( $P = 0.596$ )
Verbs, token	408.5 (142.3)	338.5 (197.0)	303.2 (182.8)	5.398	0.005	Control > aMCI ( $P = 0.095$ ), Control > naMCI ( $P = 0.007$ ), aMCI > naMCI ( $P = 0.639$ )
Noun density	17.3 (2.6)	17.2 (2.7)	17.9 (3.2)	0.723	0.487	Control > aMCI ( $P = 0.984$ ), Control < naMCI ( $P = 0.537$ ), aMCI < naMCI ( $P = 0.529$ )
Verb density	19.9 (2.0)	20.0 (2.1)	20.1 (2.3)	0.080	0.923	Control < aMCI ( $P = 0.986$ ), Control < naMCI ( $P = 0.916$ ), aMCI < naMCI ( $P = 0.976$ )
TTR	0.850 (0.02)	0.852 (0.02)	0.854 (0.02)	0.407	0.666	Control < aMCI ( $P = 0.941$ ), Control < naMCI ( $P = 0.640$ ), aMCI < naMCI ( $P = 0.871$ )

Note: Total length of talk time and token word count: 32.6 hours, 267,310 words (controls 17.8 hours, 150,864 words; aMCI 12.6 hours, 63,130 words; naMCI 11.4 hours, 53,316 words).

Abbreviations: aMCI, amnesic mild cognitive impairment; naMCI, non-amnesic mild cognitive impairment.

**TABLE 4** Means (standard deviations) of per-minute word counts and concreteness scores of all tagged words, nouns, and verbs in the speech samples of subjects with aMCI and naMCI, and of cognitively healthy controls.

	Control	aMCI	naMCI	F(2,145)	P	Post hoc Tukey contrasts
Word count						
All words	141.9 (24.3)	130.4 (25.9)	126.2 (24.2)	5.865	0.004	Control > aMCI ( $P = 0.052$ ), Control > naMCI ( $P = 0.006$ ), aMCI < naMCI ( $P = 0.751$ )
Nouns	24.3 (4.6)	22.1 (4.2)	22.4 (4.9)	3.986	0.021	Control > aMCI ( $P = 0.038$ ), Control > naMCI ( $P = 0.092$ ), aMCI < naMCI ( $P = 0.957$ )
Verbs	28.4 (6.2)	26.3 (6.7)	25.6 (6.5)	2.845	0.061	Control > aMCI ( $P = 0.232$ ), Control > naMCI ( $P = 0.078$ ), aMCI > naMCI ( $P = 0.874$ )
Concreteness score						
All words	2.55 (0.08)	2.52 (0.09)	2.56 (0.09)	1.955	0.145	Control > aMCI ( $P = 0.281$ ), Control > naMCI ( $P = 0.779$ ), aMCI < naMCI ( $P = 0.142$ )
Nouns	3.71 (0.25)	3.57 (0.28)	3.66 (0.25)	3.410	0.036	Control > aMCI ( $P = 0.027$ ), Control > naMCI ( $P = 0.682$ ), aMCI < naMCI ( $P = 0.289$ )
Verbs	2.45 (0.10)	2.43 (0.10)	2.47 (0.13)	1.176	0.311	Control > aMCI ( $P = 0.759$ ), Control < naMCI ( $P = 0.549$ ), aMCI < naMCI ( $P = 0.284$ )

Abbreviations: aMCI, amnesic mild cognitive impairment; naMCI, non-amnesic mild cognitive impairment.

literatures. In formal linguistics, nouns and verbs are recognized as the big two major lexical categories, despite the enormous cross-linguistic diversity in morphosyntactic form.<sup>29,30</sup> In cognitive neuroscience, there has been extensive evidence, from brain lesion studies to bat-

teries of word-based neuropsychological tests, that nouns and verbs are encoded in different areas of the brain, although the exact neural mechanisms in the lexical representation of grammatical categories remain a matter for debate.<sup>31–35</sup> Some patients with AD suffer from

semantic memory deficits which affect the perceptual attributes of semantic knowledge, resulting in difficulty in lexical retrieval and in more abstract speech—the reversal of the concreteness effect.<sup>1–8</sup> People with amnesic MCI also present with impairments in semantic memory, as reported in studies drawing data from common word-based neuropsychological tests, such as object and face naming tests, that target semantic memory directly.<sup>36–40</sup> Studies that draw data from connected speech, typically from picture descriptions or story recalls, are not as conclusive.<sup>14,15</sup> Our study shows that the semantic memory deficits in people with amnesic MCI are manifested in deficits in noun retrieval and in reversal of the concreteness effect involving nouns in natural speech, providing corroborative evidence that semantic memory-related deficits are manifested in ordinary language in the predementia stage of cognitive impairment. Further research is needed to substantiate these results. The incidence of MCI developing into full-fledged dementia is estimated to be between 20% and 40%.<sup>41</sup>

## 5 | CONCLUSION

To conclude, we took a corpus and formal linguistic approach to search for linguistic markers of cognitive impairment. The unconstrained natural speech data obtained from people talking about familiar topics of daily life reflect the state of the language more closely and intimately than the data collected through picture narration and fairy tale retelling, or through word-based elicitation. The choice of words and morphosyntactic phrasings in free speech is not primed, nor constrained, by pictures or fairy tales. One limitation of natural speech is that it does not shed light on the underlying pathologies of memory impairment that the semantic battery of tests, such as the Boston Naming Test,<sup>42</sup> is designed to reveal. Corpus data, however, are non-invasive and easy to collect and analyze. As we have demonstrated, an average of a little more than 10 minutes of natural talk yields adequate data that allow us to detect language deficits in people with MCI. More studies are needed, based on larger corpora of natural speech, in terms of word count and participant, to confirm the results. The corpus-linguistic method offers a reliable and cost-effective tool to detect linguistic signs of early cognitive decline, helping medical practitioners in the early diagnosis, intervention, and management of the progressive disease.

## AUTHOR CONTRIBUTIONS

ZB, LC, conceptualization, data analysis; ZB, writing, with input from all authors; KH, LL, JH, VO, and NH, corpus and data analysis; JY, TN, LF, RM, EHK, neuropsychological assessment and statistical analysis.

## ACKNOWLEDGMENTS

The authors wish to thank the Presbyterian Community Services of Singapore for supporting the study at their Hannah Senior Activity Centre at Toh Yi Drive, Singapore. The authors acknowledge the funding support of the following: Humanities and Social Sciences Research Fund, National University of Singapore; Social Science Research Thematic Grant, Ministry of Education, Singapore; DSO National Laboratories,

Singapore; Kwan Im Thong Hood Cho Temple, Singapore; Lee Im Tah Holdings, Singapore.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. The research is approved by the institutional review board of the National University of Singapore (H-17-047). Author disclosures are available in the [supporting information](#).

## CONSENT STATEMENT

All participants provided informed consent as required by the IRB proved by NUS (H-17-047). All speech data are anonymized.

## ORCID

Zhiming Bao  <https://orcid.org/0000-0002-7859-2654>

## REFERENCES

- Warrington EK. The selective impairment of semantic memory. *Q J Exp Psychol.* 1975;27(4):635–657.
- Breedin SD, Saffran EM, Coslett HB. Reversal of the concreteness effect in a patient with semantic dementia. *Cogn Neuropsychol.* 1994;11(6):617–660.
- Bird H, Lambon Ralph MA, Patterson K, Hodges JR. The rise and fall of frequency and imageability: noun and verb production in semantic dementia. *Brain Lang.* 2000;73(1):17–49.
- Bonner MF, Vesely L, Price C, et al. Reversal of the concreteness effect in semantic dementia. *Cogn Neuropsychol.* 2009;26(6):568–579.
- Maccoir J. Is a plum a memory problem? Longitudinal study of the reversal of concreteness effect in a patient with semantic dementia. *Neuropsychologia.* 2009;47(2):518–535.
- Ahmed S, Haigh AF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain.* 2013;136(12):3727–3737.
- Cousins K, Ash S, Irwin D, Grossman M. Dissociable substrates underlie the production of abstract and concrete nouns. *Brain Lang.* 2017;165:45–54.
- Cho S, Nevler N, Ash S, et al. Automated analysis of lexical features in frontotemporal degeneration. *Cortex.* 2021;137:215–231.
- Cho S, Cousins K, Shellikeri S, et al. Lexical and acoustic speech features relating to Alzheimer Disease pathology. *Neurology.* 2022;99:e313–e322.
- Hoffman P, Lambon Ralph MA. Reverse concreteness effects are not a typical feature of semantic dementia: evidence for the hub-and-spoke model of conceptual representation. *Cereb Cortex.* 2011;21(9):2103–2112.
- Hoffman P, Jones RW, Lambon Ralph MA. Be concrete to be comprehended: consistent imageability effects in semantic dementia for nouns, verbs, synonyms and associates. *Cortex.* 2013;49(5):1206–1218.
- Cuetos F, Arce N, Martínez C, Ellis AW. Word recognition in Alzheimer's disease: effects of semantic degeneration. *J Neuropsychol.* 2017;11(1):26–39.
- Peters F, Majerus S, De Baeremaeker J, Salmon E, Collette F. Impaired semantic knowledge underlies the reduced verbal short-term storage capacity in Alzheimer's disease. *Neuropsychologia.* 2009;47(14):3067–3073.
- Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: a review of picture description tasks. *J Clin Exp Neuropsychol.* 2018;40(9):917–939.
- Filiou R, Bier N, Slegers A, et al. Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: a scoping review. *Aphasiology.* 2020;34(6):723–755.

16. Lee R, Yu J, Rawtaer I, et al. CHI study: protocol for an observational cohort study on aging and mental health in community-dwelling older adults. *BMJ Open*. 2020;10(5):e035003.
17. Lim ML, Collinson SL, Feng L, Ng TP. Cross-cultural application of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): performances of elderly Chinese Singaporeans. *Clin Neuropsychol*. 2010;24(5):811-826.
18. Dugbartey AT, Townes BD, Mahurin RK. Equivalence of the color trails test and trail making test in nonnative english-speakers. *Arch Clin Neuropsychol*. 2000;15(5):425-431.
19. Maseda A, Lodeiro-Fernández L, Lorenzo-López L, et al. Verbal fluency, naming and verbal comprehension: three aspects of language as predictors of cognitive impairment. *Aging Ment Health*. 2014;18(8):1037-1045.
20. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183-194.
21. Grundman M, Petersen RC, Ferris SH, et al. Mild cognitive impairment can be distinguished from Alzheimer Disease and normal aging for clinical trials. *Arch Neurol*. 2004;61(1):59-66.
22. Csukly G, Sirály E, Fodor Z, et al. The differentiation of amnesic type MCI from the non-amnesic types by structural MRI. *Front Aging Neurosci*. 2015;8:52.
23. Bao Z. *The making of vernacular Singapore English: system, transfer and filter*. Cambridge University Press; 2015.
24. Marcus M, Santorini B, Marcinkiewicz MA. Building a large annotated corpus of English: the Penn Treebank. *Comput Linguist*. 1993;19(2):313-330.
25. Toutanova K, Klein D, Manning CD, Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. *Proc HLT-NAACL*. 2003;1:173-180.
26. Covington MA, McFall JD. Cutting the Gordian Knot: the moving-average type-token ratio (MATTR). *J Quant Linguist*. 2010;17(2):94-100.
27. Joubert S, Gardy L, Didic M, Rouleau I, Barbeau EJ. A meta-analysis of semantic memory in Mild Cognitive Impairment. *Neuropsychol Rev*. 2021;31:221-232.
28. Brysbaert M, Warriner A, Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res Methods*. 2014;46(3):904-911.
29. Croft W. *Radical construction grammar: syntactic theory in typological perspective*. Oxford University Press; 2001.
30. Evans N, Levinson SC. The myth of language universals: language diversity and its importance for cognitive science. *Behav Brain Sci*. 2009;32(5):429-492.
31. Kemmerer D. Grammatical categories. In: de Zubizaray GI, ed. *The oxford handbook of neurolinguistics*. Schiller NO. Oxford University Press; 2019.
32. Caramazza A, Hillis AE. Lexical organization of nouns and verbs in the brain. *Nature*. 1991;349(6312):788-790.
33. Shapiro K, Caramazza A. The representation of grammatical categories in the brain. *Trends Cogn Sci*. 2003;7(5):201-206.
34. Vigliocco G, Vinson DP, Druks J, Barber H, Cappa SF. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci Biobehav Rev*. 2011;35(3):407-426.
35. Kemmerer D. Word classes in the brain: implications of linguistic typology for cognitive neuroscience. *Cortex*. 2014;58:27-51.
36. Feng S, Qi R, Yang J, Yu A, Yang Y. Neural correlates for nouns and verbs in phrases during syntactic and semantic processing: an fMRI study. *J Neurolinguistics*. 2020;53:100860.
37. Adlam AR, Bozeat S, Arnold R, Watson P, Hodges JR. Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex*. 2006;42(5):675-684.
38. Ahmed S, Arnold R, Thompson SA, Graham KS, Hodges JR. Naming of objects, faces and buildings in mild cognitive impairment. *Cortex*. 2008;44(6):746-752.
39. Joubert S, Brambati SM, Ansado J, et al. The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia*. 2010;48(4):978-988.
40. Barbeau EJ, Didic M, Joubert S, et al. Extent and neural basis of semantic memory impairment in mild cognitive impairment. *J Alzheimer's Dis*. 2012;28(4):823-837.
41. Taler V, Monetta L, Sheppard C, Ohman A. Semantic function in mild cognitive impairment. *Front Psychol*. 2019;10:3041.
42. Kaplan E, Goodglass H, Weintraub S. *Boston naming test*. Lea & Febiger; 1983.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cao L, Han K, Lin L, et al. Reversal of the concreteness effect can be detected in the natural speech of older adults with amnesic, but not non-amnesic, mild cognitive impairment. *Alzheimer's Dement*. 2024;16:e12588. <https://doi.org/10.1002/dad2.12588>