# Nouns and Verbs Identify Different Subtypes of MCI

**Zhiming Bao** ( bao@nus.edu.sg )
National University of Singapore https://orcid.org/0000-0002-7859-2654
**Luwen Cao**
National University of Singapore
**Kunmei Han**
National University of Singapore
**Lin Li**
National University of Singapore
**Jia Wen Hing**
National University of Singapore
**Vincent Ooi**
National University of Singapore
**Nick Huang**
National University of Singapore
**Lei Feng**
**Rathi Mahendran**
National University of Singapore & National University Hospital & Duke-NUS medical school
**Ee Heok Kua**
National University of Singapore

**Article**

**Keywords:** semantic dementia and Alzheimer's disease, concreteness effect, mild cognitive impairment (MCI).

**Posted Date:** December 29th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1136322/v1

# Abstract

It is well-documented that patients with semantic dementia and Alzheimer's disease present with difficulty in lexical retrieval and reversal of the concreteness effect in nouns and verbs. Little is known about the lexical phenomena before the onset of symptoms. We anticipate that there are linguistic signs in the speech of people who suffer from mild cognitive impairment (MCI), the prodromal stage of dementia. Here, we report the results of a novel corpus-linguistic approach to the early detection of cognitive impairment. We recorded 40 hours of natural, unconstrained speech of 188 English-speaking Singaporeans; 90 are diagnosed with MCI (51 amnestic, 39 nonamnestic), and 98 are cognitively healthy. The recordings yield 327,470 words, which are tagged for parts of speech. We calculate the per-minute speech rates and concreteness scores of nouns and verbs, and of all tagged words, in our dataset. Our analysis shows that the two measures of nouns and verbs identify different subtypes of MCI. Compared with healthy controls, subjects with amnestic MCI produce fewer but more abstract nouns, whereas subjects with nonamnestic MCI produce fewer but more concrete verbs. Cognitive impairment is manifested in ordinary language before the presentation of clinical symptoms, and can be detected through non-invasive corpus-based analysis of natural speech.

# Main Text

In recent years, there is 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 aphasia) or Alzheimer's disease present with effortful lexical retrieval and reversal of the concreteness effect, producing nouns, and to a lesser extent, verbs, which are more abstract.[1-7] 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 or fairy tale recalls. 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 speech rate and concreteness of nouns and verbs in patients with mild cognitive impairment (MCI), the prodromal stage of dementia. In the present study, we take a novel, corpus linguistic approach to search for linguistic markers of early cognitive impairment in natural, unconstrained speech. Our analysis shows that people with MCI, especially amnestic MCI, experience lexical retrieval difficulties and reversal of the concreteness effect in nouns, consistent with the symptoms associated with semantic dementia and Alzheimer's disease.

Our language dataset comes from participants in the Community Health Intergenerational Study, a cohort study of ageing and mental health among Singaporeans 60 years of age or older.[8] Assessments include physical health, socioeconomic conditions, cognitive functioning and unconstrained speech. The neuropsychological battery of tests used to diagnose normal ageing and MCI in the study have been validated in the Singaporean population.[9] The tests evaluate amnestic and non-amnestic 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 are 1. Rey auditory verbal learning test, to evaluate declarative verbal learning and memory; 2. immediate, delayed and recognition memory test and forward and backward digit span tasks, to assess attention and verbal working memory; 3. color trail tests 1 and 2, to assess sustained attention and reasoning;[10] 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.[11] The diagnosis for each subject was arrived at through consensus review of the test results by two psychiatrists and one neuropsychologist.

Most Singaporeans are multilingual, speaking English, Chinese, Malay or Tamil. Participants were instructed to talk about any topic for up to 20 minutes in a language they felt most 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. In total, more than 475 participants provided speech samples in English. Among them, 96 were diagnosed with MCI. For our dataset, we selected 90 subjects with MCI who are in their 60s and 70s. We also selected 98 cognitively healthy participants with similar age, gender, education and language profiles. The basic demographic information of the subjects who contributed speech samples to the dataset is shown in Table 1.

Table 1: Age, gender, education and language of the subjects who volunteered to provide speech data. The means of age and education are in years, and the mean of language is in the number of languages spoken. Standard deviations are in parentheses. Df=186.

| | Control | MCI | t | *p* |
|---|---|---|---|---|
| N | 98 | 90 | | |
| Age | 66.5 (4.7) | 66.1 (5.3) | 0.49 | 0.624 |
| Gender (M/F) | 49/49 | 48/42 | | |
| Education | 14.6 (3.2) | 14.0 (3.9) | 1.11 | 0.270 |
| Language | 2.3 (0.5) | 2.2 (0.6) | 1.65 | 0.102 |

As expected, there is no significant difference between the subjects with MCI and cognitively healthy controls.

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.[12,13] 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. Table 2 displays the basic statistics of the dataset.

Table 2: Means (standard deviations) of talk time (minutes) and token counts of all words, nouns and verbs in the dataset. Total length of talk time and word count: 40 hours, 327,470 words, split between MCI (16.8 hours, 130,626 words) and healthy controls (23.2 hours, 196,844 words). Df=186 for all measures except talk time, df=167.

| | Control | MCI | t | *p* |
|---|---|---|---|---|
| N (M/F) | 49/49 | 48/42 | | |
| Talk time | 14.2 (4.2) | 11.2 (5.4) | 4.21 | <0.001 |
| All words | 2,008.6 (723.9) | 1,451.4 (838.3) | 4.89 | <0.001 |
| Nouns | 340.5 (131.3) | 246.7 (132.3) | 4.87 | <0.001 |
| Verbs | 405.1 (155.9) | 297.4 (185.6) | 4.32 | <0.001 |

As a group, people with MCI talk less, and produce fewer words, than healthy controls. The declines in the four measures are statistically significant ($p<0.001$).

For each subject's speech sample, we calculated the speech rates of nouns, verbs and all tagged words by dividing the total number of noun, verb or word 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 and colleagues, with 1 being the most abstract and 5 being the most concrete.[14] For each word or lemma, we obtained the score from the database, multiplied it with the number of tokens of the word, to arrive at the word's token score. The sum of all word token scores is then divided by the total number of word tokens to yield the concreteness score of the speech sample of each subject. About 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*).

The results are shown in Table 3.

Table 3: Means (standard deviations) of talk time (minutes), speech rates (word per minute) and concreteness scores of all words, nouns and verbs in the speech samples of healthy controls and of subjects with amnestic MCI (aMCI) and nonamnestic MCI (naMCI). For aMCI, df=147 for all measures except talk time, df=78; for naMCI, df=135 for all measures except noun concreteness score, df=57.

| | Controls | aMCI | t | *p* | naMCI | t | *p* |
|---|---|---|---|---|---|---|---|
| N (M/F) | 49/49 | 32/19 | | | 16/23 | | |
| Talk time | 14.2 (4.2) | 11.3 (5.8) | 3.25 | 0.002 | 11.2 (5.0) | 3.62 | <0.001 |
| Speech Rate | | | | | | | |
| All words | 140.5 (23.8) | 129.5 (25.9) | 2.59 | 0.011 | 122.3 (24.3) | 4.01 | <0.001 |
| Nouns | 23.7 (4.5) | 21.6 (4.5) | 2.71 | 0.007 | 22.1 (4.5) | 1.88 | 0.063 |
| Verbs | 28.3 (6.1) | 26.6 (6.9) | 1.56 | 0.120 | 24.6 (6.7) | 3.09 | 0.002 |
| Concreteness Score | | | | | | | |
| All words | 2.54 (0.08) | 2.52 (0.09) | 1.51 | 0.134 | 2.57 (0.11) | -1.44 | 0.157 |
| Nouns | 3.68 (0.24) | 3.59 (0.27) | 2.00 | 0.047 | 3.66 (0.27) | 0.38 | 0.706 |
| Verbs | 2.44 (0.11) | 2.44 (0.12) | 0.37 | 0.709 | 2.49 (0.14) | -2.07 | 0.040 |

From Table 3, we can see that the all-word measures of speech rate and concreteness score do not differentiate the two subtypes of MCI: the decline in speech rate is significant in both subtypes, and in concreteness neither the decline in amnestic MCI nor the rise in nonamnestic MCI is significant. Nouns and verbs are more discriminating. The speech rate of nouns declines in both amnestic and nonamnestic MCI, but the decline is significant in amnestic MCI, and marginally so in nonamnestic MCI ($p$=0.007 v. $p$=0.063). The speech rate of verbs, by contrast, declines in both subtypes, and the decline is significant only in nonamnestic MCI ($p$=0.120 v. $p$=0.002). The same trend is observed in concreteness. Nouns decline in concreteness, i.e. are more abstract, in both amnestic and nonamnestic MCI, but only the decline in amnestic MCI reaches the level of significance ($p$=0.047 v. $p$=0.706). Verbs maintain their concreteness score in amnestic MCI, and rise in concreteness, i.e. are more concrete, in nonamnestic MCI, and the rise is significant ($p$=0.040). It is worth emphasizing that subjects with amnestic MCI exhibited deficits in noun retrieval and reversal of the concreteness effect in nouns. Verbs are spared.

The results, and the dissociation between nouns and verbs, are consistent with the findings reported in the linguistic and neuropsychological 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.[15-17] In cognitive neuroscience, there has been extensive evidence, from brain lesion studies to batteries of word-based neuropsychological tests, that nouns and verbs are encoded in different areas in the brain, although the exact neural mechanisms in the lexical representation of grammatical categories remains a matter for debate.[18-22] Patients with semantic dementia and Alzheimer's disease 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.[1-7] People with MCI, especially amnestic 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.[23-27] Studies that draw data from connected speech, typically from picture descriptions or story recalls, are not as conclusive.[28-29] To our knowledge there has been no study on whether the semantic memory deficits in people with amnestic MCI lead to deficits in noun retrieval and in reversal of the concreteness effect. The results of our study provide the first conclusive evidence that semantic memory-related deficits are manifested in ordinary language before the presentation of clinical symptoms of full-fledged dementia.

To conclude, we took a novel, corpus linguistic approach to search for linguistic markers of cognitive impairment. The natural speech data obtained from people talking about familiar topics of daily life reflect the mental state of the language more closely and intimately than the data collected through picture narration and story re-telling, or through word-based elicitation. Moreover, corpus data are non-invasive, and easy to collect and analyze. As we have demonstrated, an average of a little more than ten minutes of natural talk yields adequate data that allow us to detect language deficits in people with mild cognitive impairment, the prodromal stage of dementia. The corpus-linguistic method reported here offers a reliable and cost-effective tool of detecting linguistic signs of cognitive decline, helping medical practitioners in the early diagnosis, intervention and management of the progressive disease.

# Methods

About Singaporean English

Singapore is a small city state of some 5.7 million people, of whom 4 million are citizens and permanent residents, according to the latest census figures released on the government's website (https://www.singstat.gov.sg/). Nearly one million residents are 60 years of age or older, constituting 22% of the population. It is a multilingual country. Since the founding of Singapore as a British crown colony in 1819, most of the early immigrants hailed from southeastern China, southern India, Malaysia and the surrounding Riau Islands of Indonesia.[29] When it gained independence in 1965, Singapore recognized four official languages, reflecting the origins of most of its immigrants: Chinese (Mandarin), Malay, Tamil, and English, with Malay having the additional title of national language, and English that of working language. For Chinese Singaporeans, in addition to Mandarin, there are mutually unintelligible dialects, the major ones being Hokkien, Teochew and Cantonese. Since the dialects share a common grammatical and lexical core,[30] we group them together as a single language. At the present time, according to the government's census survey, most Singaporeans are multilingual, with English as the dominant home language for nearly half of the households, and as the common lingua franca. Due to the constant contact with the local languages, the English language in Singapore has undergone extensive lexical and grammatical change, incorporating words (*tau* 'soy' from Chinese; *atas* 'arrogant' from Malay) and grammatical features from the local languages (*one* as a particle for emphasis).[31] Despite the fact that it is the native language of a sizable segment of the population, Singaporean English has not reached the level of register differentiation as American or British English.[32]

## The Language Samples

Our language samples are verbatim transcripts of the recordings of free-flowing speech by participants in a cohort study of ageing and mental health among Singaporeans who are 60 years of age or older. The aims and methods of the cohort study have been described elsewhere.[8] Here, we describe how language data are processed. Language sampling is entirely voluntary. The recording took place in a normal office setting, with small digital recorders. Participants were told to talk about any topic for up to 20 minutes, in a language that they felt most comfortable with. They were aware that they were being recorded. Interviewer participation was kept to the minimum to allow subjects to plan their speech as free from intervention as is practical. Words uttered by interviewers were removed from the dataset.

We used the Stanford part-of-speech tagger to assign parts of speech to the words in our dataset. The tagger uses the Penn Treebank tagset, shown below:[12,13]

Tag Description

1. CC Coordinating conjunction
2. CD Cardinal number
3. DT Determiner
4. EX Existential *there*
5. FW Foreign word
6. IN Preposition or subordinating conjunction
7. JJ Adjective
8. JJR Adjective, comparative
9. JJS Adjective, superlative

10. LS List item marker

11. MD Modal

12. NN Noun, singular or mass

13. NNS Noun, plural

14. NNP Proper noun, singular

15. NNPS Proper noun, plural

16. PDT Predeterminer

17. POS Possessive ending

18. PRP Non-possessive pronoun (personal pronoun)

19. PRP$ Possessive pronoun

20. RB Adverb

21. RBR Adverb, comparative

22. RBS Adverb, superlative

23. RP Particle

24. SYM Symbol

25. TO Infinitival *to*

26. UH Interjection

27. VB Verb, base form

28. VBD Verb, past tense

29. VBG Verb, gerund or present participle

30. VBN Verb, past participle

31. VBP Verb, non-3rd person singular present

32. VBZ Verb, 3rd person singular present

33. WDT *Wh*-determiner

34. WP *Wh*-pronoun

35. WP$ Possessive *wh*-pronoun

36. WRB *Wh*-adverb

The same tag TO is used for tagging *to* as preposition (*to school*) and as infinitival marker (*to go*). We separate the two functions, and use TO for the infinitival *to* only. We also introduced two tags, SFP for sentence-final particle (*Ok lah*) and FRG for fragments, which are common in unprepared speech (*fr-fragments*).

When tagging Singaporean English materials, the Stanford tagger's success rate is about 85%. Part of the reason for the low success rate is the frequent use of words which are unique to Singaporean English, including foreign words (*tau huay* 'soy pudding'), and English words that have developed local uses or meanings. Consider *one* as an example. It is a cardinal number (*one school*) or a pronominal (*last one*). These are tagged as *one*_CD and *one*_NN, respectively. In addition to these two uses, *one* is also used in Singaporean English as a sentence-final particle to express emphasis:

*My daughter is very active one*

'My daughter is very ACTIVE!'

The Stanford tagger tags *one* as a cardinal number here. To ensure accurate part-of-speech assignment, the tagged words were vetted by a separate group of student research assistants who were trained in formal linguistics at the National University of Singapore.

Three sample texts are shown below.

Extract 1 (64, male, cognitively healthy)

Transcribed:

I came from a a very poor family. Um I grew up in er, you know, in those days where Singapore is a slum. So I've witnessed uh riots. I've witnessed er curfew. And er then I've also experience, I've experienced when er, you know, just sharing one bowl of tau huay for 10 person in the family you know and also to the the extreme is er just er plain rice and then with a sauce and oil and sauce, you know. Sometimes this goes on for weeks ah.

Tagged:

I_PRP came_VBD from_IN a_DT a_DT very_RB poor_JJ family_NN ._. Um_UH I_PRP grew_VBD up_RP in_IN er_UH ,_, you_PRP know_VBP ,_, in_IN those_DT days_NNS where_WRB Singapore_NNP is_VBZ a_DT slum_NN ._. So_RB I_PRP 've_VBP witnessed_VBN uh_UH riots_NNS ._. I_PRP 've_VBP witnessed_VBN er_UH curfew_NN ._. And_CC er_UH then_RB I_PRP 've_VBP also_RB experience_VBP ,_, I_PRP ’ve_VBP experienced_VBN when_WRB er_UH ,_, you_PRP know_VBP ,_, just_RB sharing_VBG one_CD bowl_NN of_IN tau_FW huay_FW for_IN 10_CD person_NN in_IN the_DT family_NN you_PRP know_VBP and_CC also_RB to_IN the_DT the_DT extreme_NN is_VBZ er_UH just_RB er_UH plain_JJ rice_NN and_CC then_RB with_IN a_DT sauce_NN and_CC oil_NN and_CC sauce_NN ,_, you_PRP know_VBP ._. Sometimes_RB this_DT goes_VBZ on_RB for_IN weeks_NNS ah_SFP ._.

Extract 2 (66, male, diagnosed with amnestic MCI)

Transcribed:

why I felt that the secular practice of meditation and in in in this instance we are talking about mindfulness practice, which is really an approach, er a particular approach to meditation, can be helpful to everyone, is because er the teachings have been made secular, with hardly any reference to its religious origin, although we would mention that the approach er is founded on the b the the Buddha's teachings of meditation.

Tagged:

why_WRB I_PRP felt_VBD that_IN the_DT secular_JJ practice_NN of_IN meditation_NN and_CC in_IN in_IN in_IN this_DT instance_NN we_PRP are_VBP talking_VBG about_IN mindfulness_NN practice_NN ,_, which_WDT is_VBZ really_RB an_DT approach_NN ,_, er_UH a_DT particular_JJ approach_NN to_IN meditation_NN ,_, can_MD be_VB helpful_JJ to_IN everyone_NN ,_, is_VBZ because_IN er_UH the_DT teachings_NNS have_VBP been_VBN made_VBN secular_JJ ,_, with_IN hardly_RB any_DT reference_NN to_IN its_PRP$ religious_JJ origin_NN ,_, although_IN we_PRP would_MD mention_VB that_IN the_DT approach_NN er_UH is_VBZ founded_VBN on_IN the_DT b_FRG the_DT the_DT Buddha_NN 's_POS teachings_NNS of_IN meditation_NN ._.

Extract 3 (76, female, diagnosed with nonamnestic MCI)

Transcribed:

Erm I had my nursing training in UK, Chesterfield. During my three years there I was uh very well treated. Uh I was a happy during my training time. I enjoy my training, I have very good colleagues and nursing other other higher nursing staff. Mmm I met my husband on my third year of my training. Er we got married after my after I passed my after my finished my my training, then I work in UK for three years, then later we came home. I have I have my daughter in UK. So later on we decide to to come home to Singapore.

Tagged:

Erm_UH I_PRP had_VBD my_PRP$ nursing_NN training_NN in_IN UK_NNP ,_, Chesterfield_NNP ._. During_IN my_PRP$ three_CD years_NNS there_RB I_PRP was_VBD uh_UH very_RB well_RB treated_VBN ._. Uh_UH I_PRP was_VBD a_DT happy_JJ during_IN my_PRP$ training_NN time_NN ._. I_PRP enjoy_VBP my_PRP$ training_NN ,_, I_PRP have_VBP very_RB good_JJ colleagues_NNS and_CC nursing_NN other_JJ other_JJ higher_JJR nursing_NN staff_NN ._. Mmm_UH I_PRP met_VBD my_PRP$ husband_NN on_IN my_PRP$ third_JJ year_NN of_IN my_PRP$ training_NN ._. Er_UH we_PRP got_VBD married_VBN after_IN my_PRP$ after_IN I_PRP passed_VBD my_PRP$ after_IN my_PRP$ finished_VBD my_PRP$ my_PRP$ training_NN ,_, then_RB I_PRP work_VBP in_IN UK_NNP for_IN three_CD years_NNS ,_, then_RB later_RB we_PRP came_VBD home_RB ._. I_PRP have_VBP I_PRP have_VBP my_PRP$ daughter_NN in_IN UK_NNP ._. So_RB later_RB on_RB we_PRP decide_VBP to_TO to_TO come_VB home_RB to_IN Singapore_NNP ._.

The tagged data are processed with Antconc, a common concordance tool used by corpus linguists.[34]

Statistical Analysis

Two-tailed t-tests on age, year of education, languages spoken and talk time of the subjects who provide speech samples, and on the speech rate and concreteness score data are performed on SPSS v.27.

# Declarations

Acknowledgement

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

Contributions

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

Ethics declaration

The authors declare no competing interests.

Data Availability

The data used in the analysis are available upon request.

Code Availability

No computer code is used in the analysis of the corpus data.

Ethics declaration

The protocol was approved by the IRB of the National University of Singapore, and participants provided written, informed consent.

# References


1. Warrington, E.K. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27, 635-657 (1975).
2. Breedin, S.D., Saffran, E.M. & Coslett, H.B. Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology* 11, 617-660 (1994).
3. Bird, H., Ralph, M., Patterson, K. & Hodges, J.R. The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language* 73, 17-49 (2000).
4. Bonner, M.F. et al. Reversal of the concreteness effect in semantic dementia. *Cognitive Neuropsychology* 26, 568-579 (2009).
5. Macoir, J. Is a plum a memory problem? Longitudinal study of the reversal of concreteness effect in a patient with semantic dementia. *Neuropsychologia* 47, 518-535 (2009).
6. Cousins, K., Ash S, Irwin, D. & Grossman, M. Dissociable substrates underlie the production of abstract and concrete nouns. *Brain* & *Language* 165, 45-54 (2017).
7. Cho S. et al. Automated analysis of lexical features in frontotemporal degeneration. *Cortex* 137, 215-231 (2021).
8. Lee, R. et al. CHI study: Protocol for an observational cohort study on ageing and mental health in community-dwelling older adults. *BMJ Open* 10:e035003 (2020).
9. Dugbartey, A.T., Townes, B.D., & Mahurin, R.K. Equivalence of the color trails test and trail making test in nonnative English-speakers. *Archives of Clinical Neuropsychology* 15, 425–431 (2000).
10. Lim, M.L., Collinson, S.L., Feng, L. & Ng, T.P. Cross-cultural application of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Performances of elderly Chinese Singaporeans. *The Clinical Neuropsychologist* 811-826 (2010).
11. Maseda, A. et al. Verbal fluency, naming and verbal comprehension: Three aspects of language as predictors of cognitive impairment. *Aging Mental Health* 18, 1037–1045 (2014).
12. Marcus, M., Santorini, B. & Marcinkiewicz, M.A. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313–330 (1993).
13. Toutanova, K., Klein D., Manning C. & Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of HLT-NAACL 2003*, pp. 252-259 (2003).
14. Brysbaert, M., Warriner, A., Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res.* 46, 904-911 (2014).
15. Croft, W. *Radical construction grammar: syntactic theory in typological perspective.* Oxford University Press (2001).

16. Evans, N. & Levinson, S. C. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32, 429-492 (2009).
17. Kemmerer, D. Grammatical categories. In *The Oxford Handbook of Neurolinguistics*, ed. by de Zubicaray, G.I. & Schillar N.O. IOxford University Press, 2019).
18. Caramazza, A. & Hillis A.E. Lexical organization of nouns and verbs in the brain. *Nature* 349, 788-790 (1991).
19. Shapiro, K. & Caramazza, A. The representation of grammatical categories in the brain. *Trends in Cognitive Science* 7, 201-206 (2003).
20. Vigliocco, G. et al. Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosience and Biobehavioral Reviews* 35, 407-426 (2011).
21. Kemmerer, D. Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. *Cortex* 58, 27-51 (2014).
22. Feng, S. et al. Neural correlates for nouns and verbs in phrases during syntactic and semantic processing: An fMRI study. *Journal of Neurolinguistics* 53.100860 (2020).
23. Adlam, A.-L.R., Bozeat, S., Arnold, R., Watson, P. & Hodges, J.R. Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex* 42, 675-684 (2006).
24. Ahmed, S. et al. Naming of objects, faces and buildings in mild cognitive impairment. *Cortex* 44, 746-752 (2008).
25. Joubert, S. et al. The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia* 48, 978-988 (2010).
26. Barbeau, E.J. et al. Extent and neural basis of semantic memory impairment in Mild Cognitive Impairment. *Journal of Alzheimer's Disease* 28, 823-837 (2012).
27. Taler, V., Monetta, L., Sheppard C. & Ohman, A. Semantic function in mild cognitive impairment. *Frontiers in Psychology* 10:3041 (2019).
28. Mueller, K.D., Hermann, B., Mecollari, J. & Turkstra L.S. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology* 40, 917-939 (2018).
29. Filiou, R. et al. Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review. *Aphasiology* 34, 723-755 (2020).
30. Turnbull, C. M. *A history of Singapore 1819-1975*. (Oxford University Press, 1977).
31. Chao, Y.R. *A grammar of spoken Chinese.* (University of California Press, 1968).
32. Bao, Z. *The making of vernacular Singapore English: System, transfer and filter*. (Cambridge University Press, 2015).
33. Schneider, E.W. The dynamics of new Englishes: From identity construction to dialect birth. *Language* 79, 233-281 (2003).
34. Anthony, L. AntConc 3.5.0. Tokyo: Waseda University (2017).