Supplementary Materials

**Linguistic features extracted from samples and automatic tools used to extract them**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Linguistic domain** | **Features** | **Explanatory notes** | **Extraction tool** |
| **Lexico-syntactic (275)** | Word production & complexity (11) | 1. DESWLsy: mean number of syllables per word 2. DESWLsyd: sd of the mean number of syllables per word 3. DESWLlt: mean number of characters per word 4. DESWLltd: sd of the mean number of characters per word 5. WRDPOLc: mean polysemy for content words 6. WRDHYPn: mean hypernymy for nouns 7. WRDHYPv: mean hypernymy for verbs 8. WRDHYPnv: mean hypernymy for nouns and verbs 9. Sixltr: % words >6 letters 10. repeated\_content\_lemmas: % content words repeated in sample, lemmatized 11. repeated\_content\_and\_pronoun\_lemmas: % content words and pronouns repeated in sample, lemmatized | Polysemy is the number of different senses of a word, e.g.to *book* a ticket or read a *book.*  Hypernymy is calculated according to the number of words superordinate to the target word in a taxonomic hierarchy (Graesser, Namara, Louwerse, & Cai, 2004). Both are calculated using the WordNet database. | 1-8 Coh-Metrix 3.0 (Graesser et al., 2004)  9 LIWC2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015)  10 & 11 TAACO 2.0.4 (Crossley, Kyle, & Dascalu, 2019) |
| Parts-of-speech (POS) (18) | 1. nouns: % Nouns 2. verbs: % Verbs 3. inflected\_verbs: % Inflected verbs 4. light: % Light verbs 5. function: % Function words 6. pronouns: % Pronouns 7. determiners: % Determiners 8. adverbs: % Adverbs 9. adjectives: % Adjectives 10. prepositions: % Prepositions 11. coordinate: % Coordinate conjunctions 12. subordinate: % Subordinate conjunctions 13. demonstratives: % Demonstratives 14. nvratio: Noun:verb ratio (nouns / verbs) 15. noun\_ratio: Noun ratio (nouns / (nouns + verbs)) 16. prp\_ratio: Pronoun ratio (pronouns / (pronouns + nouns)) 17. sub\_coord\_ratio: Subordinate:coordinate ratio (subordinate conjunctions / coordinate conjunctions) 18. NID: ‘not in dictionary’, % words that do not appear in the English dictionary | Features are part-of-speech counts (normalized by number of words) and ratios. | COVFEFE (Liaqat, Fraser, & Komeili, 2019), utilizes the Stanford POS tagger |
| Lexical richness (8) | 1. TTR: Type-token-ratio or *U/V*, where *U* is the number of unique words (types) and *V* is the total words used (tokens) 2. brunet: Brunét’s index or *V(U-0.165)* 3. honore: Honoré’s statistic or ﻿100log*N*/(1-(*N*1/*U*)) where *N*1 is the number of words used only once (hapax legomena). ﻿Calculates hapax legomena as a proportion of the total number of words used as an indication of richness. 4. MATTR\_10: Moving average TTR (window size = 10) 5. MATTR\_20: Moving average TTR (window size = 20) 6. MATTR\_30: Moving average TTR (window size = 30) 7. MATTR\_40: Moving average TTR (window size = 40) 8. MATTR\_50: Moving average TTR (window size = 50) | TTR is a measure of word re-use, indicating lexical richness. All features use content words only. Features 2-8 are approaches to measuring lexical richness that aim to avoid issues associated with samples of different lengths (Covington & McFall, 2010). | COVFEFE (Liaqat et al., 2019) |
| Psycholinguistics (34) | 1. MRC\_Familiarity\_CW: mean familiarity rating per content word, Medical Research Council (MRC) norms 2. MRC\_Familiarity\_FW: as above for function words 3. MRC\_Familiarity\_AW: as above for all words 4. MRC\_Concreteness\_CW: mean concreteness rating per content word, MRC norms 5. MRC\_Concreteness\_FW: as above for function words 6. MRC\_Concreteness\_AW: as above for all words 7. MRC\_Imageability\_CW: mean imageability rating per content word, MRC norms 8. MRC\_Imageability\_FW: as above for function words 9. MRC\_Imageability\_AW: as above for all words 10. MRC\_Meaningfulness\_CW: mean meaningfulness rating per content word, MRC norms 11. MRC\_Meaningfulness\_FW: as above for function words 12. MRC\_Meaningfulness\_AW: as above for all words 13. Kuperman\_AoA\_CW: mean age-of-acquisition rating per content word, Kuperman norms 14. Kuperman\_AoA\_FW: as above for function words 15. Kuperman\_AoA\_AW: as above for all words 16. Brysbaert\_Concreteness\_Combined\_CW: mean concreteness rating per content word, Brysbaert norms 17. Brysbaert\_Concreteness\_Combined\_FW: as above for function words 18. Brysbaert\_Concreteness\_Combined\_AW: as above for all words 19. SUBTLEXus\_Freq\_CW: mean frequency rating per content word, SUBTL norms from the SUBTLEXus corpus 20. SUBTLEXus\_Freq\_FW: as above for function words 21. SUBTLEXus\_Freq\_AW: as above for all words 22. SUBTLEXus\_Range\_CW: mean range per content word according to SUBTLEXus corpus. 23. SUBTLEXus\_Range\_FW: as above for function words 24. SUBTLEXus\_Range\_AW: as above for all words 25. BNC\_Spoken\_Freq\_CW: mean frequency rating per content word, British National Corpus (BNC) 2007 (spoken) corpus 26. BNC\_Spoken\_Freq\_FW: as above for function words 27. BNC\_Spoken\_Freq\_AW: as above for all words 28. BNC\_Spoken\_Range\_CW: mean range per content word according to BNC 2007 (spoken) corpus 29. BNC\_Spoken\_Range\_FW: as above for function words 30. BNC\_Spoken\_Range\_AW: as above for all words 31. BNC\_Spoken\_Bigram\_Normed\_Freq: mean frequency for bigrams according to BNC 2007 (spoken) corpus 32. BNC\_Spoken\_Bigram\_Proportion: proportion of bigrams in transcript that are within the most frequent 50,000 bigrams of the BNC 2007 (spoken) corpus 33. BNC\_Spoken\_Trigram\_Normed\_Freq: mean frequency for trigrams according to BNC 2007 (spoken) corpus 34. BNC\_Spoken\_Trigram\_Proportion: proportion of trigrams in transcript that are within the most frequent 50,000 bigrams of the BNC 2007 (spoken) corpus | Psycholinguistics indicate lexical sophistication by measuring different phenomena associated with words in the lexicon, and are assembled from different huma ratings. Familiar words are more known to users of the language. Concrete words are associated with the five senses, i.e. we can hear/ see/ feel/ smell or touch it, as opposed to abstract words. Imageability is the degree to which an image of the word can be created.  Meaningfulness measures how related a word is to other words.  Age-of-acquisition is the age at which language learners are usually exposed to a word.  Frequency count features use a large reference corpus, e.g. BNC Consortium 2007, to estimate the frequency of each word. The SUBTLEXus corpus calculates word usage across subtitles from film and television (Brysbaert & New, 2009).  Not all words will have an associated score for each measure, and these words are excluded from the total count when averaging.  Range indices are a different approach to measuring frequency. The range of a word is the number of documents it appears in, i.e. how widely a word is used (Kyle & Crossley, 2015). | TAALES 2.2 (Kyle & Crossley, 2015; Kyle, Crossley, & Berger, 2018) |
| Psychological processes (50) | 1. Analytic: reflecting academic/ analytical thinking (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014) 2. Clout: reflecting social standing or ‘rank’ (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014) 3. Authentic: reflecting truth (Newman, Pennebaker, Berry, & Richards, 2003) 4. Tone: reflecting emotional tone (Cohn, Mehl, & Pennebaker, 2004) 5. affect e.g. happy, cried 6. posemo: positive emotion e.g. love, sweet 7. negamo: negative emotion e.g. hurt, nasty 8. anx: anxiety e.g. worried 9. anger e.g. hate, kill 10. sad e.g. crying 11. social: social processes e.g. mate, talk 12. family e.g. daughter, dad 13. friend e.g. buddy 14. female: female references e.g. girl, her 15. male: male references e.g. boy, his 16. cogproc: cognitive processes e.g. know 17. insight e.g. think, know 18. cause: causation e.g. because, effect 19. discrep: discrepancy e.g. should, would 20. tentat: tentative e.g. maybe, perhaps 21. certain: certainty e.g. always, never 22. differ: differentiation e.g. but, else 23. percept: perceptual processes e.g. look, heard 24. see e.g. view, saw 25. hear e.g. listen, hearing 26. feel e.g. feels, touch 27. bio: biological processes e.g. eat, pain 28. body e.g. cheeks, hands 29. health e.g. clinic, flu 30. sexual e.g. love, incest 31. ingest e.g. dish, pizza 32. drives 33. affiliation e.g. ally, friend 34. achieve: achievement e.g. win, success 35. power e.g. superior, bully 36. reward e.g. take, prize 37. risk e.g. danger, doubt 38. focuspast: words focused on the past e.g. ago, did 39. focuspresent: words focused on the present e.g. today, now 40. focusfuture: words focused on the future e.g. may, soon 41. relativ: relativity e.g. area, bend 42. motion e.g. arrive, go 43. space e.g. down, in 44. time e.g. end, season 45. work e.g. job, majors 46. leisure e.g. cook, chat 47. home e.g. kitchen 48. money e.g. cash, owe 49. relig: religion e.g. alter, church 50. death e.g. bury, coffin | 1-4 are summary linguistic variables derived from previous research from the LIWC lab and converted to a % score.  5-50 are the % words in the sample relating to psychological constructs and personal concerns according to LIWC2015 internal dictionary of word categories. The same word can appear in multiple categories. Definitions supplied where required. | LIWC2015 (Pennebaker et al., 2015) |
| Syntactic structures & complexity (32) | 1. WordCount: Total number of words 2. S: Total number of sentences 3. VP: Number of verb phrases normalized by total words (verb phrases/ words) 4. C: Number of clauses normalized by total words (clauses / words). 5. T: Number of T-units normalized by total words (T-units/ words). 6. DC: Number of dependent clauses normalized by total words (dependent clauses/ words). 7. CP: Number of coordinate phrases normalized by total words (coordinate phrases/ words) 8. CN: Number of complex nominals normalized by total words (complex nominals/ words) 9. CT\_A: Clauses per T-unit   **Length of production:**   1. MLC: Mean length of clause (﻿words / clauses) 2. MLS: Mean length of sentence (﻿words / sentences) 3. MLT: Mean length of T-unit (﻿words / T-units)   **Sentence complexity:**   1. CS: Sentence complexity ratio (clauses per sentence or ﻿clauses/sentences)   **Amount of subordination:**   1. CT: T-unit complexity ratio (clauses per T-unit) 2. CTT: Complex T-unit ratio (complex T-units per T-unit) 3. DCC: Dependent clause ratio (dependent clauses per clause) 4. DCT: Dependent clauses per T-unit   **Amount of coordination:**   1. CPC: Coordinate phrases per clause 2. CPT: Coordinate phrases per T-unit 3. TS: T-units per sentence   ﻿**Relationship between syntactic structures and larger production units:**   1. CNC: Complex nominals per clause 2. CNT: Complex nominals per T-unit 3. VPT: Verb phrases per T-unit 4. PP\_type\_rate: rate of prepositional phrases (PP/words) 5. PP\_type\_prop: proportion of prepositional phrases (PP length/words) 6. VP\_type\_rate: rate of verb phrases (VP/words) 7. VP\_type\_prop: proportion of verb phrases (VP length/words) 8. NP\_type\_rate: rate of noun phrases (NP/words) 9. NP\_type\_prop: proportion of noun phrases (NP length/words) 10. average\_PP\_length: words in prepositional phrases/ total prepositional phrases across sample 11. average\_VP\_length: as above for verb phrases 12. average\_NP\_length: as above for noun phrases | Clauses correspond to a subject and predicate, and all its modifiers.  A T-unit corresponds to a main clause and all attached dependent clauses. A complex T-unit contains a dependent clause.  A dependent clause contains a subject and a verb but cannot constitute a sentence alone.  Complex nominals occur when a head noun is preceded by a modifier.  Headings in bold correspond to five categories determined by Lu (2010). | COVFEFE (utilizes Lu’s Syntactic Complexity Analyzer (SCA, (Lu, 2010)) for features 1 – 23) |
| Syntactic parse tree features (4) | 1. maxdepth: Maximum Yngve depth of each parse tree, averaged over all sentences 2. totaldepth: Total sum of the Yngve depths for each parse tree, averaged over all sentences 3. meandepth: Mean Yngve depth of the parse tree, averaged over all sentences 4. treeheight: Average height of parse trees across all sentences | Calculations based on the Yngve depth for each word in a top-down, left-to-right syntax tree of a parsed sentence. It is a measure of embeddedness, or the ‘stack’ at each word (Yngve, 1960). The score increases as embeddedness increases, and features are different calculations based on the score. | COVFEFE (Liaqat et al., 2019), utilizes the Stanford parser |
| Grammatical constituents of syntax tree (111) | 1. ROOT\_gt\_S 2. NP\_gt\_DT\_NN 3. PP\_gt\_IN\_NP 4. S\_gt\_NP\_VP 5. NP\_gt\_PRP 6. NP\_gt\_NNS 7. S\_gt\_NP\_VP\_A 8. S\_gt\_VP 9. PRT\_gt\_RP 10. ADVP\_gt\_RB 11. NP\_gt\_NP\_PP 12. NP\_gt\_DT\_NNS 13. S\_gt\_CC\_NP\_VP 14. VP\_gt\_VBZ\_VP 15. NP\_gt\_NN 16. ROOT\_gt\_NP 17. SBAR\_gt\_S 18. VP\_gt\_TO\_VP 19. NP\_gt\_DT 20. S\_gt\_VP\_A 21. NP\_gt\_DT\_NN\_NN 22. NP\_gt\_DT\_JJ\_NN 23. VP\_gt\_VBZ\_NP 24. VP\_gt\_VB\_NP 25. VP\_gt\_VBG\_NP 26. ROOT\_gt\_FRAG 27. NP\_gt\_PRP$\_NN 28. VP\_gt\_VBP\_NP 29. VP\_gt\_VB 30. NP\_gt\_FW 31. NP\_gt\_NP\_VP 32. SBAR\_gt\_IN\_S 33. NP\_gt\_NP\_CC\_NP 34. ADJP\_gt\_JJ 35. VP\_gt\_VBG\_PP 36. NP\_gt\_JJ\_NN 37. VP\_gt\_VBP 38. NP\_gt\_NP\_SBAR 39. NP\_gt\_NP\_VP\_A 40. VP\_gt\_VBP\_SBAR 41. S\_gt\_ADVP\_NP\_VP 42. VP\_gt\_VBP\_S 43. SBAR\_gt\_WHNP\_S 44. VP\_gt\_MD\_VP 45. PP\_gt\_TO\_NP 46. VP\_gt\_VP\_CC\_VP 47. VP\_gt\_VBZ 48. VP\_gt\_VBZ\_S 49. VP\_gt\_VBG 50. VP\_gt\_VBG\_S 51. WHNP\_gt\_WP 52. NP\_gt\_NNP 53. VP\_gt\_VBP\_VP 54. VP\_gt\_VBG\_PRT 55. FRAG\_gt\_CC\_NP 56. NP\_gt\_NP\_SBAR\_A 57. VP\_gt\_VB\_S 58. S\_gt\_S\_CC\_S 59. VP\_gt\_VBZ\_PP 60. ADVP\_gt\_RB\_RB 61. VP\_gt\_VBG\_PRT\_PP 62. VP\_gt\_VBZ\_ADJP 63. NP\_gt\_NP\_NP 64. NP\_gt\_RB 65. NP\_gt\_NN\_NN 66. VP\_gt\_VBD\_NP 67. WHNP\_gt\_WDT 68. NP\_gt\_NN\_NNS 69. VP\_gt\_VB\_NP\_PP 70. NP\_gt\_NP\_PP\_A 71. NP\_gt\_PRP$\_NNS 72. NP\_gt\_NP\_NP\_A 73. NP\_gt\_EX 74. INTJ\_gt\_UH 75. NP\_gt\_DT\_JJ\_NNS 76. S\_gt\_INTJ\_VP 77. VP\_gt\_VB\_VP 78. NP\_gt\_CD\_NNS 79. VP\_gt\_VBG\_NP\_PP 80. VP\_gt\_VBD\_SBAR 81. VP\_gt\_VB\_PP 82. VP\_gt\_VBN\_PP 83. ADJP\_gt\_RB\_JJ 84. VP\_gt\_VBZ\_SBAR 85. WHADVP\_gt\_WRB 86. FRAG\_gt\_ADJP 87. SBAR\_gt\_WHADVP\_S 88. VP\_gt\_VBP\_PP 89. S\_gt\_NP\_ADJP 90. S\_gt\_NP\_ADVP\_VP 91. NP\_gt\_DT\_DT\_NN 92. PP\_gt\_IN 93. VP\_gt\_VBD\_VP 94. NP\_gt\_CD 95. VP\_gt\_VBN\_NP 96. S\_gt\_NP\_NP 97. PP\_gt\_IN\_PP 98. ROOT\_gt\_INTJ 99. ROOT\_gt\_SBARQ 100. S\_gt\_CC\_NP\_VP\_A 101. VP\_gt\_VBG\_PRT\_NP 102. VP\_gt\_VB\_ADJP 103. VP\_gt\_VBZ\_NP\_PP 104. NP\_gt\_PRP$\_JJ\_NN 105. NP\_gt\_DT\_FW 106. NP\_gt\_JJ\_NNS 107. ADJP\_gt\_JJ\_PP 108. ADVP\_gt\_RB\_PP 109. VP\_gt\_VB\_SBAR 110. PP\_gt\_IN\_S 111. NP\_gt\_CD\_NN | Constituents comprising the syntactic parse tree are quantified and normalized by the total number of constituents in the sample.  For example NP\_gt\_DT\_NN represents a noun phrase (NP) comprised of a determiner (DT) and noun (NN), usually denoted NP -> DT NN.  Tags are from the Penn Treebank, see <https://web.archive.org/web/20130517134339/http://bulba.sdsu.edu/jeanette/thesis/PennTags.html> for a full list of clause and phrase level tags. Word level tags CC - Coordinating conjunction CD - Cardinal number DT - Determiner EX - Existential there FW - Foreign word IN - Preposition or subordinating conjunction JJ - Adjective JJR - Adjective, comparative JJS - Adjective, superlative LS - List item marker MD - Modal NN - Noun, singular or mass NNS - Noun, plural NNP - Proper noun, singular NNPS - Proper noun, plural PDT - Predeterminer POS - Possessive ending PRP - Personal pronoun PRP$ - Possessive pronoun (prolog version PRP-S) RB - Adverb RBR - Adverb, comparative RBS - Adverb, superlative RP - Particle SYM - Symbol TO - to UH - Interjection VB - Verb, base form VBD - Verb, past tense VBG - Verb, gerund or present participle VBN - Verb, past participle VBP - Verb, non-3rd person singular present VBZ - Verb, 3rd person singular present WDT - Wh-determiner WP - Wh-pronoun WP$ - Possessive wh-pronoun (prolog version WP-S) WRB - Wh-adverb | COVFEFE (Liaqat et al., 2019) |
| Shannon entropy (1) | 1. Entropy: H 2 ( X ) = − ∑ i = 1 n c o u n t i N log 2 ⁡ ( c o u n t i N ) Entropy for letters, given below, where *N* is total letters and *counti* is the count of letter*i*, | Entropy, arising from information theory and applied here to letters, is a measure of information inherent in the sample. It indicates certainty with which an unknown letter can be predicted based on previous known information (Shannon, 1951). | Python script\* |
| Fluency (3) | 1. False\_starts\_ratio: false starts/total outputs 2. Filler\_ratio: fillers/total outputs 3. Nonspecifics\_ratio: ‘thing’ words/words | False starts correspond to words that are started but not completed, e.g. “he looks **nau=** naughty”.  Fillers correspond to ‘um’, ‘ah’, ‘er’ and their variations.  Non-specific ‘thing’ words are thing, some*thing*, any*thing* | Python script\* |
| Non-verbal (3) | 1. Laughter: Count of laughter normalized by ‘total outputs’ i.e. words + laughter + pauses (laughter/total outputs) 2. Pauses\_ratio: As above for pauses (pauses/total outputs) 3. QMark: Count of the number of questions asked according to instances of ‘?’ | Laughter and pauses were annotated at the transcription stage by the transcriber. Duration was not recorded. | 1 & 2 Python script\*  3 LIWC2015 (Pennebaker et al., 2015) |
| **Semantic (11)** | Semantic content (3) | 1. Idea\_density: Propositional idea density 2. prop\_density: Similar to above, calculates how dense a sample is with propositions using the following -(verbs+adjectives+adverbs+prepositions+conjunctions)/ words 3. content\_density: Calculates how dense with content bearing words a sample is using - (nouns+verbs+adjectives+adverbs)/words | Idea density is a measure of how dense a sample is with ‘ideas’, based on POS tags. CPIDR 3.2 propositions roughly correspond to ﻿verbs, adjectives, adverbs, prepositions and conjunctions. Certain rules are applied to make adjustments prior to the calculation stage and has been found to closely match human ratings (Brown, Snodgrass, Kemper, Herman, & Covington, 2008) | 1 CPIDR 3.2 (Brown et al., 2008)  2 & 3 (Liaqat et al., 2019) |
| Semantic coherence (8) | 1. Mean\_cosine 2. Mean\_cosine\_1 3. sd\_cosine\_1 4. Mean\_cosine\_2 5. sd\_cosine\_2 6. Mean\_cosine\_4 7. sd\_cosine\_4 8. Mean\_wmd: average Word Movers Distance between adjacent sentences | Features utilize semantic space word vector representations to explore semantic coherence. Vectors are pre-trained word2vec model using the Google News corpus (Mikolov, Chen, Corrado, & Dean, 2013). Features 1-7 calculate mean cosine similarity between adjacent sentences and moving windows, and standard deviation (sd) of the mean as a measure of variability.  Word Movers Distance (WMD) calculates the minimum cumulative distance needed to travel between word vectors of adjacent sentences as a measure of similarity (Kusner, Sun, Kolkin, & Weinberger, 2015). | Python script\* |

\*Python scripts are available at https://github.com/natasha-clarke/CCD-Study.git

**Full results for HC vs. AD+MCI classification using different connected speech tasks and univariate or multivariate feature selection approaches**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **PD** | **CS** | **ONR** | **PR** | **NNR** |
| **All features** | Accuracy | 0.78 (0.13) | 0.68 (0.08) | 0.80 (0.12) | 0.60 (0.19) | 0.58 (0.13) |
| Sensitivity | 0.69 (0.30) | 0.62 (0.29) | 0.71 (0.31) | 0.66 (0.23) | 0.67 (0.23) |
| Specificity | 0.85 (0.15) | 0.80 (0.20) | 0.90 (0.15) | 0.59 (0.30) | 0.58 (0.35) |
| AUC | 0.83 (0.07) | 0.72 (0.14) | 0.85 (0.11) | 0.76 (0.27) | 0.67 (0.13) |
| ***k*=5** | Accuracy | 0.58 (0.11) | 0.56 (0.11) | 0.68 (0.15) | 0.54 (0.15) | 0.58 (0.18) |
| Sensitivity | 0.46 (0.09) | 0.56 (0.11) | 0.64 (0.17) | 0.58 (0.29) | 0.50 (0.26) |
| Specificity | 0.72 (0.24) | 0.60 (0.20) | 0.75 (0.27) | 0.55 (0.33) | 0.68 (0.11) |
| AUC | 0.73 (0.12) | 0.70 (0.13) | 0.75 (0.16) | 0.63 (0.21) | 0.65 (0.13) |
| ***k*=10** | Accuracy | 0.66 (0.13) | **0.66 (0.11)** | 0.72 (0.13) | 0.50 (0.19) | **0.62 (0.16)** |
| Sensitivity | 0.60 (0.30) | **0.62 (0.10)** | 0.68 (0.22) | 0.56 (0.31) | **0.53 (0.21)** |
| Specificity | 0.76 (0.16) | **0.78 (0.31)** | 0.79 (0.15) | 0.47 (0.34) | **0.72 (0.11)** |
| AUC | 0.70 (0.14) | **0.74 (0.10)** | 0.81 (0.10) | 0.67 (0.29) | **0.62 (0.10)** |
| ***k*=20** | Accuracy | 0.60 (0.10) | 0.62 (0.13) | 0.62 (0.25) | 0.58 (0.24) | 0.58 (0.19) |
| Sensitivity | 0.58 (0.16) | 0.53 (0.30) | 0.71 (0.09) | 0.65 (0.27) | 0.57 (0.29) |
| Specificity | 0.57 (0.19) | 0.79 (0.29) | 0.61 (0.38) | 0.53 (0.39) | 0.58 (0.32) |
| AUC | 0.69 (0.15) | 0.66 (0.19) | 0.69 (0.26) | 0.71 (0.25) | 0.60 (0.37) |
| ***k*=40** | Accuracy | 0.72 (0.15) | 0.56 (0.09) | 0.66 (0.17) | 0.68 (0.19) | 0.46 (0.11) |
| Sensitivity | 0.73 (0.22) | 0.42 (0.25) | 0.53 (0.28) | 0.74 (0.20) | 0.47 (0.33) |
| Specificity | 0.64 (0.20) | 0.79 (0.29) | 0.79 (0.24) | 0.61 (0.42) | 0.48 (0.16) |
| AUC | 0.77 (0.12) | 0.70 (0.10) | 0.73 (0.22) | 0.80 (0.25) | 0.54 (0.22) |
| **10RFE-LogR** | Accuracy | **0.76 (0.18)** | 0.60 (0.14) | **0.78 (0.08)** | **0.74 (0.15)** | 0.56 (0.09) |
| Sensitivity | **0.69 (0.30)** | 0.67 (0.16) | **0.75 (0.23)** | **0.78 (0.15)** | 0.56 (0.15) |
| Specificity | **0.81 (0.12)** | 0.51 (0.32) | **0.82 (0.21)** | **0.74 (0.25)** | 0.58 (0.14) |
| AUC | **0.84 (0.11)** | 0.68 (0.12) | **0.84 (0.05)** | **0.85 (0.19)** | 0.66 (0.17) |

10RFE-LogR = Recursive feature elimination (with a step size of 1) combined with logistic regression to select 10 features. PD = picture description, CS = conversational speech, ONR = overlearned narrative, PR = procedural recall, NNR = novel narrative recall. Results in bold are highest accuracies using feature selection (and associated sensitivity, specificity and AUC) for each speech task, presented in the main paper.

**Full results for HC vs AD classification using different connected speech tasks and univariate or multivariate feature selection approaches**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **PD** | **CS** | **ONR** | **PR** | **NNR** |
| **All features** | Balanced accuracy | 0.59 (0.24) | 0.72 (0.12) | 0.77 (0.18) | 0.52 (0.20) | 0.69 (0.18) |
| Sensitivity | 0.40 (0.42) | 0.52 (0.15) | 0.58 (0.28) | 0.38 (0.26) | 0.63 (0.22) |
| Specificity | 0.78 (0.33) | 0.92 (0.12) | 0.95 (0.11) | 0.66 (0.23) | 0.75 (0.20 |
| AUC | 0.83 (0.23) | 0.86 (0.03) | 0.83 (0.18) | 0.65 (0.30) | 0.82 (0.17) |
| ***k*=5** | Balanced accuracy | 0.50 (0.19) | 0.62 (0.09) | 0.83 (0.15) | **0.68 (0.24)** | **0.70 (0.28)** |
| AUC | 0.58 (0.32) | 0.63 (0.19) | 0.91 (0.09) | **0.65 (0.25)** | **0.76 (0.27)** |
| Sensitivity | 0.33 (0.31) | 0.35 (0.22) | 0.75 (0.35) | **0.52 (0.46)** | **0.60 (0.42)** |
| Specificity | 0.67 (0.42) | 0.89 (0.15) | 0.91 (0.12) | **0.84 (0.15)** | **0.80 (0.15)** |
| ***k*=10** | Balanced accuracy | 0.47 (0.25) | 0.69 (0.20) | 0.85 (0.16) | 0.41 (0.09) | 0.71 (0.18) |
| Sensitivity | 0.30 (0.45) | 0.47 (0.36) | 0.75 (0.35) | 0.22 (0.22) | 0.65 (0.34) |
| Specificity | 0.63 (0.28) | 0.91 (0.12) | 0.95 (0.11) | 0.60 (0.14) | 0.76 (0.22) |
| AUC | 0.59 (0.36) | 0.80 (0.21) | 0.96 (0.06) | 0.46 (0.16) | 0.73 (0.26) |
| ***k*=20** | Balanced accuracy | 0.58 (0.25) | **0.75 (0.15)** | **0.90 (0.11)** | 0.42 (0.20) | 0.48 (0.20) |
| Sensitivity | 0.40 (0.42) | **0.62 (0.26)** | **0.83 (0.24)** | 0.27 (0.25) | 0.38 (0.26) |
| Specificity | 0.77 (0.22) | **0.88 (0.12)** | **0.96 (0.09)** | 0.56 (0.29) | 0.58 (0.25) |
| AUC | 0.61 (0.26) | **0.80 (0.23)** | **0.94 (0.06)** | 0.40 (0.31) | 0.65 (0.21) |
| ***k*=40** | Balanced accuracy | 0.48 (0.15) | 0.65 (0.21) | 0.74 (0.07) | 0.38 (0.18) | 0.67 (0.20) |
| Sensitivity | 0.33 (0.31) | 0.42 (0.28) | 0.53 (0.07) | 0.30 (0.45) | 0.68 (0.32) |
| Specificity | 0.63 (0.28) | 0.88 (0.16) | 0.95 (0.11) | 0.46 (0.15) | 0.65 (0.21) |
| AUC | 0.55 (0.24) | 0.64 (0.08) | 0.79 (0.13) | 0.40 (0.27) | 0.77 (0.21) |
| **10RFE-LogR** | Balanced accuracy | **0.59 (0.30)** | 0.58 (0.27) | 0.69 (0.13) | 0.55 (0.26) | 0.62 (0.15) |
| Sensitivity | **0.50 (0.35)** | 0.42 (0.28) | 0.43 (0.25) | 0.45 (0.37) | 0.52 (0.29) |
| Specificity | **0.68 (0.32)** | 0.75 (0.28) | 0.95 (0.11) | 0.65 (0.31) | 0.73 (0.22) |
| AUC | **0.75 (0.26)** | 0.68 (0.32) | 0.77 (0.14) | 0.65 (0.20) | 0.78 (0.24) |

10RFE-LogR = Recursive feature elimination (with a step size of 1) combined with logistic regression to select 10 features. PD = picture description, CS = conversational speech, ONR = overlearned narrative, PR = procedural recall, NNR = novel narrative recall. Results in bold are highest accuracies using feature selection (and associated sensitivity, specificity and AUC) for each speech task, presented in the main paper.

**Full results for HC vs MCI classification using different connected speech tasks and univariate or multivariate feature selection approaches**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **PD** | **CS** | **ONR** | **PR** | **NNR** |
| **All features** | Balanced accuracy | 0.70 (0.21) | 0.62 (0.16) | 0.74 (0.15) | 0.63 (0.22) | 0.49 (0.20) |
| Sensitivity | 0.65 (0.42) | 0.60 (0.42) | 0.50 (0.35) | 0.47 (0.36) | 0.32 (0.21) |
| Specificity | 0.75 (0.15) | 0.65 (0.34) | 0.97(0.06) | 0.79 (0.14) | 0.67 (0.28) |
| AUC | 0.74 (0.24) | 0.62 (0.15) | 0.88 (0.22) | 0.70 (0.38) | 0.35 (0.21) |
| ***k*=5** | Balanced accuracy | 0.44 (0.20) | 0.57 (0.21) | 0.56 (0.17) | 0.31 (0.14) | 0.38 (0.16) |
| Sensitivity | 0.17 (0.24) | 0.27 (0.25) | 0.35 (0.42) | 0.00 (0.00) | 0.17 (0.24) |
| Specificity | 0.71 (0.28) | 0.88 (0.17) | 0.76 (0.16) | 0.63 (0.29) | 0.59 (0.30) |
| AUC | 0.46 (0.29) | 0.70 (0.35) | 0.70 (0.29) | 0.26 (0.08) | 0.27 (0.11) |
| ***k*=10** | Balanced accuracy | 0.57 (0.13) | 0.43 (0.25) | 0.70 (0.15) | 0.47 (0.23) | 0.33 (0.15) |
| Sensitivity | 0.37 (0.41) | 0.15 (0.22) | 0.52 (0.29) | 0.27 (0.43) | 0.07 (0.15) |
| Specificity | 0.78 (0.24) | 0.70 (0.31) | 0.88 (0.17) | 0.66 (0.24) | 0.60 (0.20) |
| AUC | 0.68 (0.21) | 0.51 (0.21) | 0.68 (0.17) | 0.50 (0.15) | 0.27 (0.20) |
| ***k*=20** | Balanced accuracy | 0.59 (0.17) | **0.70 (0.20)** | 0.67 (0.14) | 0.54 (0.20) | **0.43 (0.15)** |
| Sensitivity | 0.37 (0.41) | **0.58 (0.37)** | 0.47 (0.36) | 0.27 (0.43) | **0.30 (0.27)** |
| Specificity | 0.81 (0.19) | **0.82 (0.19)** | 0.88 (0.17) | 0.66 (0.24) | **0.56 (0.17)** |
| AUC | 0.67 (0.11) | **0.75 (0.10)** | 0.72 (0.20) | 0.49 (0.28) | **0.50 (0.23)** |
| ***k*=40** | Balanced accuracy | **0.62 (0.26)** | 0.56 (0.14) | 0.75 (0.20) | 0.44 (0.21) | 0.40 (0.12) |
| Sensitivity | **0.40 (0.42)** | 0.37 (0.22) | 0.62 (0.36) | 0.30 (0.45) | 0.27 (0.25) |
| Specificity | **0.84 (0.15)** | 0.76 (0.14) | 0.88 (0.17) | 0.59 (0.23) | 0.53 (0.23) |
| AUC | **0.77 (0.28)** | 0.63 (0.07) | 0.70 (0.28) | 0.47 (0.18) | 0.42 (0.16) |
| **10RFE-LogR** | Balanced accuracy | 0.59 (0.10) | 0.47 (0.20) | **0.78 (0.13)** | **0.52 (0.12)** | 0.35 (0.12) |
| Sensitivity | 0.37 (0.22) | 0.28 (0.39) | **0.67 (0.31)** | **0.43 (0.25)** | 0.00 (0.00) |
| Specificity | 0.80 (0.19) | 0.65 (0.31) | **0.90 (0.10)** | **0.60 (0.19)** | 0.70 (0.24) |
| AUC | 0.71 (0.22) | 0.53 (0.23) | **0.82 (0.22)** | **0.62 (0.21)** | 0.42 (0.25) |

10RFE-LogR = Recursive feature elimination (with a step size of 1) combined with logistic regression to select 10 features. PD = picture description, CS = conversational speech, ONR = overlearned narrative, PR = procedural recall, NNR = novel narrative recall. Results in bold are highest accuracies using feature selection (and associated sensitivity, specificity and AUC) for each speech task, presented in the main paper.

**Supplementary Materials References**

﻿BNC Consortium. (2007), British National Corpus, version 3 (BNC XML ed.). Retrieved from <http://www.natcorp.ox.ac.uk>

﻿Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of prepositional idea density from part-of-speech tagging. *Behavior Research Methods*, *40*(2), 540–545. https://doi.org/10.3758/BRM.40.2.540

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100. https://doi.org/10.1080/09296171003643098

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, *51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Graesser, A. C., Namara, D. S. M. C., Louwerse, M. A. X. M., & Cai, Z. (2004). Coh-Metrix : Analysis of text on cohesion and language, *36*(2), 193–202.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Kusner, M. J., Sun, Y., Kolkin, N. I., & Weinberger, K. Q. (2015). From Word Embeddings To Document Distances. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 957–966).

Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, *49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, *50*(3), 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Liaqat, D., Fraser, K. C., & Komeili, M. (2019). Talk2Me : Automated linguistic data collection for personal assessment. *PLoS ONE*, 1–24.

Lu, X. (2010). *Automatic analysis of syntactic complexity in second language writing*. *International Journal of Corpus Linguistics* (Vol. 15). https://doi.org/10.1075/ijcl.15.4.02lu

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space, 1–12. https://doi.org/10.1162/153244303322533223

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Technical Journal*, *30*(1), 50–64. https://doi.org/10.1002/j.1538-7305.1951.tb01366.x

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The Development and Psychometric Properties of LIWC2015. *UT Faculty/Researcher Works*, 1–22. https://doi.org/10.15781/T29G6Z

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, *9*(12). <https://doi.org/10.1371/journal.pone.0115844>

Yngve,V. 1960. A model and hypothesis for language structure. Proceedings of the American Physical Society, 104:444–466.