SCOPE: The South Carolina psycholinguistic metabase

Chuanji Gao[1] · Svetlana V. Shinkareva[2] · Rutvik H. Desai[2]

Abstract

The number of databases that provide various measurements of lexical properties for psycholinguistic research has increased rapidly in recent years. The proliferation of lexical variables, and the multitude of associated databases, makes the choice, comparison, and standardization of these variables in psycholinguistic research increasingly difficult. Here, we introduce The South Carolina Psycholinguistic Metabase (SCOPE), which is a metabase (or a meta-database) containing an extensive, curated collection of psycholinguistic variable values from major databases. The metabase currently contains 245 lexical variables, organized into seven major categories: General (e.g., frequency), Orthographic (e.g., bigram frequency), Phonological (e.g., phonological uniqueness point), Orth-Phon (e.g., consistency), Semantic (e.g., concreteness), Morphological (e.g., number of morphemes), and Response variables (e.g., lexical decision latency). We hope that SCOPE will become a valuable resource for researchers in psycholinguistics and affiliated disciplines such as cognitive neuroscience of language, computational linguistics, and communication disorders. The availability and ease of use of the metabase with comprehensive set of variables can facilitate the understanding of the unique contribution of each of the variables to word processing, and that of interactions between variables, as well as new insights and development of improved models and theories of word processing. It can also help standardize practice in psycholinguistics. We demonstrate use of the metabase by measuring relationships between variables in multiple ways and testing their individual contribution towards a number of dependent measures, in the most comprehensive analysis of this kind to date. The metabase is freely available at go.sc.edu/scope.

Keywords Psycholinguistic · Database · Lexical characteristics · Word recognition

How words are processed is a central question in psycholinguistics and cognitive science in general. A rich body of research has addressed this question using behavioral studies, most commonly using lexical decision and word naming tasks. This has provided valuable insights into properties that affect word processing, shedding light on mechanisms behind both visual and auditory word processing. In the last 15 years, large-scale studies ("megastudies"), often including tens of thousands of words, have greatly boosted this research. For example, the English Lexicon Project (ELP) provides measures of both visual lexical decision and speeded naming tasks for over 40,000 words, along with numerous psycholinguistic properties (Balota et al., 2007). Other megastudies provide data for visual lexical decision of British English (Keuleers et al., 2012), auditory lexical decision (Goh et al., 2020; Tucker et al., 2019), and semantic decision (Pexman et al., 2017). Instead of relying on a small sample of carefully chosen words that may be idiosyncratic in some fashion, these studies enable development and testing of potentially more robust word processing models on a larger scale. At the same time, the number of studies that measure word properties by collecting ratings on psycholinguistic variables for thousands of words has also increased rapidly (Taylor et al., 2020), providing rich and robust measures of many lexical properties (e.g., Brysbaert et al., 2014; Lynott et al., 2020; Pexman et al., 2019; Scott et al., 2019).

A welcome result of this work is an extensive and ever-increasing list of psycholinguistic variables that are found to affect, to varying extents, how words are processed. When investigating the contribution of a particular variable, researchers control for effects of other variables that

✉ Svetlana V. Shinkareva
  shinkareva@sc.edu

✉ Rutvik H. Desai
  rutvik@sc.edu

1   Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

2   Department of Psychology, Institute for Mind and Brain, University of South Carolina, Columbia, SC 29201, USA

are deemed to be standard control variables. However, what is considered "standard" often differs significantly between studies. Typically, a handful of variables such as frequency, number of letters or phonemes, and imageability or concreteness are used, in addition to a few other variables that vary. However, the number of psycholinguistic variables known to affect word processing is much larger. The justification for using a particular set of variables is often not clear. Investigators' research interest and expertise are understandably important factors in the choice, as well as convenience or ease of obtaining values for these variables. With the proliferation of megastudies that provide ratings on myriad of variables, it is increasingly difficult to identify and obtain values for all potentially relevant variables. This is because these variables are often spread across many datasets, most of which provide ratings for partially different sets of words. Some databases also contain values for multiple senses of the same word forms (e.g., Scott et al., 2019).

Moreover, in some cases, a nominally single variable has several subtlety different versions. For example, word frequency has well over a dozen measures. Many other variables, such as affective ratings or imageability also have multiple measures available. The contribution of different versions of a variable to a dependent variable is likely different (Brysbaert & New, 2009). Investigators often use a particular version of a variable that is familiar, or customary in their lab. For example, one lab may often use the log_Freq_HAL variable from the ELP as frequency measure as a convention or habit, while another may use log_SUBTLEX as standard practice. Ideally, this choice should be driven by a systematic comparison between different versions taking into account factors such as variances explained by a particular version of a variable, differences of tasks instructions, goals of the study, and dependent measures of interest.

To overcome these challenges in psycholinguistic research, we introduce a new meta-database or metabase named SCOPE (South CarOlina Psycholingusitic mEtabase). It aims to provide the most comprehensive collection of variables to date, by integrating megastudies and other major databases in the form of a curated metabase. It also contains additional variables for both words and nonwords not found in other databases. We have attempted to be extensive in our coverage, while acknowledging that the number of possible lexical variables is virtually unlimited. No database can be truly comprehensive, because new variables are being proposed and measured all the time, but we hope to update the metabase periodically to include new variables. Our hope is that this metabase will be, for many cases, a "one stop shop" for psycholinguistics and for affiliated disciplines such as cognitive neuroscience of language, computational linguistics, and communication disorders. Ease of use and availability may enhance incorporation of many of the

provided variables in psycholinguistic and neurolinguistic studies, potentially leading to faster progress.

Our aims are threefold. First, we expect that the metabase will enable a more comprehensive examination of word processing, which will result in a better understanding of the unique contribution of each of the variables and their interactions. It may promote the development of improved models and theories of word and language processing, and a better exchange of insights from different facets of language research, such as reading, speech perception, and semantics. Second, we hope that such a metabase will help standardize practice in psycholinguistics and will lead to a wider agreement over which variables, and which versions of those variables, are the most informative and should be routinely used or controlled for in different contexts. Finally, we present a preliminary analysis that measures the relationship between a large subset of the variables, and their individual contributions to several dependent variables, in the most comprehensive analysis of this type to date.

The variables in the metabase are organized in seven groups. The first group, "General", contains variables that have elements of orthography, phonology, and semantics, and are often strongly predictive of word processing performance. Variables such as frequency and age-of-acquisition are contained in this group. Three groups correspond to major components of a lexical item: "Orthographic", "Phonological" and "Semantic". A fifth group, "Orth-Phon", contains variables that represent the relationship between orthography and phonology. A sixth group is "Morphological," which contains variables such as length of morpheme and morpheme frequency. Finally, the seventh group contains "Response" variables represented by mean response times and accuracies. In addition to words, the database also contains orthographic and phonological measures as well as dependent measures, when available, for some pseudowords and nonwords.

We describe the distribution of each of the variables, and the relationship between each of the independent variables and the dependent variables. The metabase contains 245 variables (some of which are multi-dimensional) and a total of 105,992 words and 81,934 nonwords in the current version, with varying number of variable values available for each item. Finally, with each variable, we provide associated information such as the definition and the citation of its source that should be used when that variable is included in a study, as it is essential to credit the original creators of these databases.

Description of database

The variables in the SCOPE metabase (Supplemental Table 1; Supplemental material can be found at https://osf.io/9qbjz/) were divided into General, Orthographic,

Phonological, Semantic, Orth-Phon, Morphological, and Response Variable groups. Each variable is briefly described below.

General variables

Freq_HAL

Log10 version of frequency norms based on the Hyperspace Analogue to Language (HAL) corpus (Lund & Burgess, 1996). It contains text from approximately 3000 Usenet newsgroups that is very conversional and noisy, like spoken language.

Freq_KF

Log10 version of frequency norms based on the Kucera and Francis corpus (Kučera & Francis, 1967).

Freq_SUBTLEXUS

Log10 version of frequency norms based on the SUBTLEXus corpus (Brysbaert & New, 2009). The main sources of SUBTLEXus corpus are American television and film subtitles.

Freq_SUBTLEXUS_Zipf

A standardized version of Freq_SUBTLEXUS that can be interpreted independently of the corpus size, and it is calculated based on the equation: Freq_SUBTLEXUS_Zipf $= \log10(\frac{\text{frequency count}+1}{\text{size of corpus}+\text{num of word types}}) + 3$, in which the size of corpus and number of word types are in millions. It is a standardized measure with the same interpretation irrespective corpus size, and it has multiple advantages compared with frequency per million words (Brysbaert & New, 2009; Van Heuven et al., 2014).

Freq_SUBTLEXUK

Log10 version of the frequency norms based on SUBTLEXuk corpus (Van Heuven et al., 2014). Like Freq_SUBTLEXUS, it is based on British film and television subtitles rather than books and other written sources.

Freq_SUBTLEXUK_Zipf

A standardized version of Freq_SUBTLEXUK that can be interpreted independently of the corpus size and it is computed in the same way as Freq_SUBTLEXUS_Zipf, for which Freq_SUBTLEXUK_Zipf $= \log10(\frac{\text{frequency count}+1}{\text{size of corpus}+\text{num of word types}}) + 3$ (Van Heuven et al., 2014).

Freq_Blog

Log10 version of the frequency norms based on sources from blogs (Gimenes & New, 2016).

Freq_Twitter

Log10 version of the frequency norms based on sources from Twitter (Gimenes & New, 2016).

Freq_News

Log10 version of the frequency norms based on sources from newspapers (Gimenes & New, 2016).

Freq_Cob

Log10 of word frequencies in English based on COBUILD corpus (Baayen et al., 1996).

Freq_CobW

Log10 of word frequencies in written English based on COBUILD corpus (Baayen et al., 1996).

Freq_CobS

Log10 of word frequencies in spoken English based on COBUILD corpus (Baayen et al., 1996).

Freq_Cob_Lemmas

Log10 of lemma frequencies in English based on COBUILD corpus (Baayen et al., 1996), which is the sum of frequencies of all the infected forms of a particular word.

Freq_CobW_Lemmas

Log10 of lemma frequencies in written English based on COBUILD corpus (Baayen et al., 1996).

Freq_CobS_Lemmas

Log10 of lemma frequencies in spoken English based on COBUILD corpus (Baayen et al., 1996).

CD_SUBTLEXUS

Log10 version of the contextual diversity of a word, which refers to the number of passages in the SUBTLEXus corpus containing a particular word (Brysbaert & New, 2009).

CD_SUBTLEXUK

Log10 version of the contextual diversity of a word based on the SUBTLEXuk corpus (Van Heuven et al., 2014).

CD_Blog

Log10 version of the contextual diversity of a word based on blog sources (Gimenes & New, 2016).

CD_Twitter

Log10 version of the contextual diversity of a word based on Twitter sources (Gimenes & New, 2016).

CD_News

Log10 version of the contextual diversity of a word based on news sources (Gimenes & New, 2016).

Fam_Glasgow

A word's rated subjective familiarity on a 1 (unfamiliar) to 7 (familiar) scale (Scott et al., 2019).

Fam_Brys

Percentage of participants who know the word well enough to rate its concreteness (Brysbaert et al., 2014).

Prevalence_Brys

The proportion of participants who know the word. Participants were asked to indicate whether they knew the stimulus (yes or no) in a list of words and nonwords. Percentages were translated to *z* values based on cumulative normal distribution. A word known by 2.5% of the participants corresponds to a word prevalence of –1.96; a word known by 97.5% of the participants corresponds to a prevalence of +1.96 (Brysbaert et al., 2019).

AoA_Kuper

The age at which people acquired the word. Participants were asked to enter the age (in years) at which they estimated they had learned the word (Kuperman et al., 2012).

AoA_LWV

The age at which people acquired the word, in which a three-choice test was administered to participants in grades 4 to 16 (college) (Living Word Vocabulary database) (Dale & O'Rourke, 1981).

AoA_Glasgow

Rated age of acquisition, which indicates the age at which people estimate they acquired the word on 1 (early) to 7 (late) scale (Scott et al., 2019).

Freqtraj_TASA

How experience with a word is distributed over time based on TASA corpus. It was computed by first taking logarithms of the frequencies and then transforming them to *z*-values for low (grades 1–3) and high grades (grades 11–13), respectively (Brysbaert, 2017).

Cumfreq_TASA

Total amount of exposure to a word across time based on TASA corpus. It was computed by first taking logarithms of the frequencies at different grade levels from grade 1 to 13,

transforming them to *z*-values and then obtaining the sum of the *z*-values (Brysbaert, 2017).

DPoS_Brys

The dominant grammatical category to which a word is assigned in accordance with its syntactic functions (Brysbaert et al., 2012).

DPoS_VanH

The dominant grammatical category to which a word is assigned in accordance with its syntactic functions (Van Heuven et al., 2014).

SCOPE_ID

Unique ID for each word. This was chosen to be the same as the ELP ID (Balota et al., 2007) for words/nonwords that are in the ELP database, and new values were created for other items.

ELP_ID

Unique ID for each word from the ELP database (Balota et al., 2007), when available.

Orthographic variables

NLett

Number of letters in a word.

UnigramF_Avg_C
UnigramF_Avg_C_Log

The average frequency of the constrained unigrams of a word and its log10 version. A constrained unigram is defined as a specific letter in a specific position, for words of a specific length (Medler & Binder, 2005).

UnigramF_Avg_U
UnigramF_Avg_U_Log

The average frequency of the unconstrained unigrams for a word and its log10 version. An unconstrained unigram is defined as a specific letter within a word, regardless of its position, or the word length (Medler & Binder, 2005).

BigramF_Avg_C
BigramF_Avg_C_Log

The average frequency of the constrained bigrams for a word and its log10 version. A constrained bigram is defined as a specific two letter combination (bigram) within a word, in a specific position, for words of a specific length (Medler & Binder, 2005).

BigramF_Avg_U
BigramF_Avg_U_Log

The average frequency of the unconstrained bigrams for a word and its log10 version. An unconstrained bigram is defined as a specific two letter combination (bigram) within a word, regardless of its position, or word length (Medler & Binder, 2005).

TrigramF_Avg_C
TrigramF_Avg_C_Log

The average frequency of the constrained trigrams for a word and its log10 version. A constrained trigram is defined as a specific three letter combination (trigram) in a specific position, for words of a specific length (Medler & Binder, 2005).

TrigramF_Avg_U
TrigramF_Avg_U_Log

The average frequency of the constrained trigrams for a word. An unconstrained trigram is defined as a specific three letter combination (trigram) within a word, regardless of its position, or the word length (Medler & Binder, 2005).

OLD20

The Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood defined as the mean Levenshtein distance of a word to its 20 closest orthographic neighbors (Balota et al., 2007; Yarkoni et al., 2008). Levenshtein Distance is the minimum number of substitution, insertion or deletion operations required to change one word to another (Levenshtein, 1966).

OLD20F

The mean log HAL frequency of the closest 20 Levenshtein Distance orthographic neighbors (Balota et al., 2007; Yarkoni et al., 2008).

Orth_N

Orthographic neighborhood (or Coltheart's N), which is the number of words that can be obtained by changing one letter while preserving the identity and positions of the other letters (Balota et al., 2007; Coltheart, 1977).

Orth_N_Freq

The average frequency of the orthographic neighborhood of a particular word (Balota et al., 2007).

Orth_N_Freq_G

The number of words in the orthographic neighborhood of an item with a frequency greater than the frequency of the item (Balota et al., 2007).

Orth_N_Freq_G_Mean

The average frequency of the orthographic neighbors who have a frequency greater than the given word (Balota et al., 2007).

Orth_N_Freq_L

The number of orthographic neighbors with a frequency less than that of a given item (Balota et al., 2007).

Orth_N_Freq_L_Mean

The average frequency of the orthographic neighbors who have a frequency lower than the given word (Balota et al., 2007).

Orth_Spread

The number of letter positions that can be changed to form a neighbor that differs by a single letter (Chee et al., 2020).

OUP

Orthographic uniqueness point of a word. It indicates which letter position within the word distinguishes it from all other words. An index of one greater than the number of letters is assigned if even the final letter of the word does not make it unique (Tucker et al., 2019; Weide, 2005).

Phonological variables

NPhon

The number of phonemes in a word (Balota et al., 2007).

NSyll

The number of syllables in a word (Balota et al., 2007).

UniphonP_Un
UniphonP_St

The average likelihood of each phoneme occurring in each position of a word weighted by SUBTLEXus frequency, with vowel-stress ignored (Un) or distinct stress-vowels considered (St) (Vaden et al., 2009). It was calculated by averaging the positional probabilities of the constituent phonemes of a word in their respective positions (i.e., frequency of each phoneme occurring in a specific position).

UniphonP_Un_C
UniphonP_St_C

The length-constrained average likelihood of each phoneme occurring in each position of a word weighted by SUBTLEXus frequency, with vowel-stress ignored (Un) or distinct stress-vowels considered (St) (Vaden et al., 2009).

BiphonP_Un
BiphonP_St

The relative frequency of the sound sequences of a word at the level of its phoneme pairs (i.e., number of items a phoneme pair occurs among all words divided by all pairwise counts) weighted by SUBTLEXus frequency, with vowel-stress ignored (Un) or stress-vowels accounted for (St) (Vaden et al., 2009; Vitevitch & Luce, 1999).

TriphonP_Un
TriphonP_St

The relative frequency of the sound sequences of a word at the level of its phoneme triplets weighted by SUBTLEXus frequency, with vowel-stress ignored (Un) or distinct stress placement distinguished (St) (Vaden et al., 2009).

PLD20

Phonological Levenshtein Distance 20, which is the mean Levenshtein distance of a word to its 20 closest phonological neighbors (Balota et al., 2007; Yarkoni et al., 2008).

PLD20F

The mean log frequency of the closest 20 Levenshtein distance phonological neighbors (Balota et al., 2007; Yarkoni et al., 2008).

Phon_N

Phonological neighborhood, measured by the number of words that can be obtained by changing one phoneme while preserving the identity and positions of the other phonemes (Balota et al., 2007).

Phon_N_Freq

The average logHAL frequency of the phonological neighborhood of a particular word (Balota et al., 2007).

Phon_Spread

The number of phoneme positions that can be changed to form a neighbor that differs by a single phoneme (Chee et al., 2020).

PUP

Phonological uniqueness point of a word based on CMU Pronouncing Dictionary, which indicates which phoneme position within the word distinguishes it from all other words. An index of one greater than the number of letters is assigned if even the final sound of the word does not make it unique (Tucker et al., 2019; Weide, 2005).

Phon_Cluster_Coef

The fraction of neighbors of a word that are also phonological neighbors of each other (Goldstein & Vitevitch, 2014).

First_Phon

The first phoneme of a word based on the CMU Pronouncing Dictionary (Weide, 2005). It includes the following 14 dimensions coded with a binary code: bilabial, labiodental, interdental, alveolar, palatal, velar, glottal, stop, fricative, affricate, nasal, liquid, glide, and voiced.

IPA Transcription

Phonemic transcription using the International Phonetic Alphabet, based on the CMU Pronouncing Dictionary (Weide, 2005).

Semantic variables

Conc_Brys

The degree to which the concept can be experienced directly through the senses on a 1 (abstract) to 5 (concrete) scale (Brysbaert et al., 2014).

Conc_Glasgow

The degree to which the concept can be experienced directly through the senses on a 1 (abstract) to 7 (concrete) scale (Scott et al., 2019).

Imag_Glasgow

The degree of effort involved in generating a mental image of something on a 1 (unimageable) to 7 (imageable) scale (Scott et al., 2019).

Imag_Composite

The degree of effort involved in generating a mental image of a concept on a scale from 1 (unimageable) to 7 (imageable) (see Graves et al., 2010 for details). This measure was obtained from a database compiled from six sources, and ratings of words present in multiple databases were averaged (Bird et al., 2001; Clark & Paivio, 2004; Cortese & Fugett, 2004; Gilhooly & Logie, 1980; Paivio et al., 1968; Toglia & Battig, 1978).

Nsenses_WordNet

Number of senses based on the WordNet database (Miller, 1995). A sense is a discrete representation of one aspect of the meaning of a word.

Nsenses_Wordsmyth

Number of senses based on the Wordsmyth dictionary (Rice et al., 2019).

Nmeanings_Wordsmyth

Number of meanings based on the Wordsmyth dictionary (Rice et al., 2019).

Nmeanings_Websters

Number of meanings based on the Websters dictionary, which was computed in the current paper by counting the number of distinct entries under the same wordform presented in the Websters dictionary.

NFeatures

Number of features listed for a word (Buchanan et al., 2019). This measure was obtained by asking participants to provide lists of features for each concept presented.

Visual_Lanc

To what extent one experiences the referent by seeing, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Auditory_Lanc

To what extent one experiences the referent by hearing, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Haptic_Lanc

To what extent one experiences the referent by feeling through touch, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Olfactory_Lanc

To what extent one experiences the referent by smelling, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Gustatory_Lanc

To what extent one experiences the referent by tasting, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Interoceptive_Lanc

To what extent one experiences the referent by sensations inside one's body, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Head_Lanc

To what extent one experiences the referent by performing an action with the head, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Torso_Lanc

To what extent one experiences the referent by performing an action with the torso, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Mouth_Throat_Lanc

To what extent one experiences the referent by performing an action with the mouth/throat, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Hand_Arm_Lanc

To what extent one experiences the referent by performing an action with the hand/arm, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Foot_Leg_Lanc

To what extent one experiences the referent by performing an action with the foot/leg, from 0 (not experienced at all) to 5 (experienced greatly) (Lynott et al., 2020).

Mink_Perceptual_Lanc

Minkowski distance at $m = 3$ of an 11-dimension sensorimotor vector from the origin. It represents a composite measure of the perceptual strength in all dimensions, with the influence of weaker dimensions attenuated (Lynott et al., 2020).

Mink_Action_Lanc

Minkowski distance at $m = 3$ of an 11-dimension sensorimotor vector from the origin. It represents a composite measure of the action strength in all dimensions, with the influence of weaker dimensions attenuated (Lynott et al., 2020).

Compo_attribs [65]

Componential Attributes: A set of 65 experiential attributes based on neurobiological considerations, comprising sensory, motor, spatial, temporal, affective, social, and cognitive experiences on a 0 (not at all) to 6 (very much) scale (Binder et al., 2016).

BOI

Body–Object Interaction, which is the ease with which the human body can interact with a word's referent on a scale from 1 (low interaction) to 7 (high interaction) (Pexman et al., 2019).

Sem_Size_Glasgow

Magnitude of an object or concept expressed in either concrete (physical) or abstract terms on a 1 (small) to 7 (big) scale (Scott et al., 2019).

Gender_Assoc_Glasgow

The degree to which words are associated with male or female behavior on a 1 (feminine) to 7 (masculine) scale (Scott et al., 2019).

Feature_Visual

The word is associated with sense of vision as indicated by 0 or 1 rated by two English speakers; disagreements were discussed and agreed upon (Vinson & Vigliocco, 2008).

Feature_Perceptual

The word describes information gained through sensory input, including body state and proprioception as indicated by 0 or 1 rated by two English speakers; disagreements were discussed and agreed upon (Vinson & Vigliocco, 2008).

Feature_Functional

The word refers to the purpose of a thing, or the purpose or goal of an action as indicated by 0 or 1 rated by two English speakers; disagreements were discussed and agreed upon (Vinson & Vigliocco, 2008).

Feature_Motoric

The word describes a motor component of an action as indicated by 0 or 1 rated by two English speakers; disagreements were discussed and agreed upon (Vinson & Vigliocco, 2008).

Sensory_Experience

The extent to which a word evokes a sensory and/or perceptual experience in the mind of the reader on a 1 to 7 scale, with higher numbers indicating a greater sensory experience (Juhasz & Yap, 2013).

Socialness

The extent to which a word's meaning has social relevance on a seven-point Likert scale from 1 (not social) to 7 (highly social) (Diveica et al., 2022).

Valence_Warr

The pleasantness of a stimulus on a 1 (happy) to 9 (unhappy) scale (Warriner et al., 2013).

Valence_Extremity_Warr

The absolute value of the difference between valence rating from 5, the neutral point on the scale (Warriner et al., 2013).

Valence_Glasgow

The pleasantness of a stimulus on a 1 (happy) to 9 (unhappy) scale (Scott et al., 2019).

Valence_NRC

Word-emotion association built by manual annotation using Best-Worst Scaling method, with scores ranging from 0 (negative) to 1 (positive) (Mohammad & Turney, 2010; Mohammad & Turney, 2013).

Arousal_Warr

The intensity of emotion evoked by a stimulus on a 1 (aroused) to 9 (calm) scale (Warriner et al., 2013).

Arousal_Glasgow

The intensity of emotion evoked by a stimulus on a 1 (aroused) to 9 (calm) scale (Scott et al., 2019).

Arousal_NRC

Word-emotion association built by manual annotation using Best-Worst Scaling method, with scores ranging from 0 (low arousal) to 1 (high arousal) (Mohammad & Turney, 2010; Mohammad & Turney, 2013).

Dominance_Warr

The degree of control exerted by a stimulus on a 1 (controlled) to 9 (in control) scale (Warriner et al., 2013).

Dominance_Glasgow

The degree of control exerted by a stimulus on a 1 (controlled) to 9 (in control) scale (Scott et al., 2019).

Dominance_NRC

Word-emotion association built by manual annotation using Best-Worst Scaling method, with scores ranging from 0 (low dominance) to 1 (high dominance) (Mohammad & Turney, 2010; Mohammad & Turney, 2013).

Humor_Male_Enge
Humor_Female_Enge
Humor_Young_Enge
Humor_Old_Enge
Humor_Overall_Enge

Humor ratings on a scale from 1 (humorless) to 5 (humorous) for group of raters that are male/female or young/old, and an overall rating (Engelthaler & Hills, 2018).

Emot_Assoc [10]

Word-emotion association built by manual annotation, with 0 (not associated) and 1 (associated) ratings for 10 emotions: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust (Mohammad & Turney, 2010; Mohammad & Turney, 2013).

Sem_Diversity

The degree to which different contexts associated with a word vary in their meanings. In other words, similarity of different contexts in which a word can appear. This is a computationally derived measure of semantic ambiguity, which is more objective compared to the measure by summing the number of senses or dictionary definitions (Hoffman et al., 2013).

Sem_N

The number of semantic neighbors within a threshold determined in a co-occurrence space. The space was created from a sparse matrix that contains all co-occurrence information for each word with window size and weighting scheme applied. The threshold is calculated by randomly sampling many word pairs and calculating their interword distances to obtain the mean and standard deviation of this distance distribution (Shaoul & Westbury, 2006, 2010). The threshold was 1.5 SDs below the mean distance.

Sem_N_D

The average radius of co-occurrence, which is the average distance between the words in the semantic neighborhood and the target word (Shaoul & Westbury, 2006, 2010).

Sem_N_D_Taxonomic_N3
Sem_N_D_Taxonomic_N10
Sem_N_D_Taxonomic_N25
Sem_N_D_Taxonomic_N50

The mean distance of nearest 3, 10, 25, or 50 semantic neighbors of a word based on taxonomic similarity. Similarity is calculated using vector representations (calculated from a corpus) of words that emphasize taxonomic (as opposed to thematic or associative) relations (Reilly & Desai, 2017; Roller & Erk, 2016).

Assoc_Freq_Token

The number of times that a word is the first associate across all target words. The task instruction was to elicit free associations in the broadest possible sense wherein participants were asked to provide multiple responses per cue (De Deyne et al., 2019).

Assoc_Freq_Type

The number of unique words that produce the target word first in a free association task (De Deyne et al., 2019).

Assoc_Freq_Token123

The number of times that a word is one of the first three associates across all target words in a free association task (De Deyne et al., 2019).

Assoc_Freq_Type123

The number of unique words that produce the target word in the first three associates in a free association task (De Deyne et al., 2019).

Cue_SetSize

The number of different responses or targets given by two or more participants in the normative sample, which provides a relative index of the set size of a particular word by providing a reliable measure of how many strong associates it has (Nelson et al., 2004).

Cue_MeanConn

The number of connections among the associate set of a word, divided by the size of the set, which captures the density and in some sense the level of organization among the strongest associates of the cue (Nelson et al., 2004).

Cue_Prob

The probability that each associate in a set produces the normed cue as an associate (Nelson et al., 2004).

Cue_ResoStrength

Resonance strength between the cue and its associates, calculated by cross-multiplying cue-to-associate strength by associate-to-cue strength for each associate in a set and then summing the result (Nelson et al., 2004).

Word2Vec [300]

Vector representation of a word created from 300 hidden layer linear units in the neural net model trained on the Google news dataset (Mikolov et al., 2013).

GloVe [300]

Vector representation of a word created from an unsupervised learning algorithm. Training is performed on aggregated global word-word co-occurrence statistics from a corpus (Pennington et al., 2014).

Taxonomic [300]

Vector representation of a word created from a model that uses a narrow window of co-occurrence, effectively emphasizing taxonomic similarities between words as opposed to associations (Roller & Erk, 2016). In most distributional models, words such as *cow* and *milk* have similar representations due to their high association, while the distance between *cow* and *bull* is relatively greater. These representations reverse this relationship, and assign a greater similarity to *cow* and *bull*.

Orth-Phon variables

Phonographic_N

The number of words that can be obtained by changing one letter and one phoneme while preserving the identity and position of the other letters and phonemes (Balota et al., 2007; Peereman & Content, 1997).

Phonographic_N_Freq

The average frequency of the phonographic neighborhood of the particular word (Balota et al., 2007).

Consistency_Token_FF

The spelling-to-sound consistency measure, in which a given word's log frequencies of friends are divided by its total log frequencies of friends and enemies. In addition to the composite value, this measure also includes token feedforward onset (Consistency_Token_FF_O), nucleus (Consistency_Token_FF_N), coda (Consistency_Token_FF_C), oncleus (Consistency_Token_FF_ON), and rime (Consistency_Token_FF_R) consistency (Chee et al., 2020).

Consistency_Token_FB

The sound-to-spelling consistency measure, in which a given word's log frequencies of friends are divided by its total log frequencies of friends and enemies. In addition to the composite value, this measure also includes token feedback onset (Consistency_Token_FB_O), nucleus (Consistency_Token_FB_N), coda (Consistency_Token_FB_C), oncleus (Consistency_Token_FB_ON), and rime (Consistency_Token_FB_R) consistency (Chee et al., 2020).

Consistency_Type_FF

The spelling-to-sound consistency measure, in which a given word's number of friends were divided by its total number of friends and enemies. In addition to the composite value, this measure also includes type feedforward onset (Consistency_Type_FF_O), nucleus (Consistency_Type_FF_N), coda (Consistency_Type_FF_C), oncleus (Consistency_Type_FF_ON), and rime (Consistency_Type_FF_R) consistency (Chee et al., 2020).

Consistency_Type_FB

The sound-to-spelling consistency measure, in which a given word's number of friends were divided by its total number of friends and enemies. In addition to the composite value, this measure also includes type feedback onset (Consistency_Type_FB_O), nucleus (Consistency_Type_FB_N), coda (Consistency_Type_FB_C), oncleus (Consistency_Type_FB_ON), and rime (Consistency_Type_FB_R) consistency (Chee et al., 2020).

Morphological variables

NMorph

The number of morphemes in a word (Sánchez-Gutiérrez et al., 2018).

PRS_signature

A prefix-root-suffix signature (Sánchez-Gutiérrez et al., 2018). For example, words that include one suffix and one root, but no prefix, share a 0-1-1 PRS signature.

ROOT1_Freq_HAL
ROOT2_Freq_HAL
ROOT3_Freq_HAL

The summed frequency of all members in the morphological family of a morpheme occurring as the first (ROOT1), second (ROOT2), or third (ROOT3) root (Sánchez-Gutiérrez et al., 2018).

SUFF1_Freq_HAL
SUFF2_Freq_HAL
SUFF3_Freq_HAL
SUFF4_Freq_HAL

The summed frequency of all members in the morphological family of a morpheme occurring as the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4) suffix (Sánchez-Gutiérrez et al., 2018).

ROOT1_FamSize
ROOT2_FamSize
ROOT3_FamSize

The number of word types in which a given morpheme is a constituent as the first (ROOT1), second (ROOT2), or third (ROOT3) root. It was computed by counting all its types in the ELP database (Sánchez-Gutiérrez et al., 2018).

SUFF1_FamSize
SUFF2_FamSize
SUFF3_FamSize
SUFF4_FamSize

The number of word types in which a given morpheme is a constituent as the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4) suffix. It was computed by counting all its types in the ELP database (Sánchez-Gutiérrez et al., 2018).

ROOT1_PFMF
ROOT2_PFMF
ROOT3_PFMF

Percentage of other words in the family that are more frequent for the first (ROOT1), second (ROOT2), or third (ROOT3) root. It was computed by dividing the number of

more frequent words in the family by the total number of members in the family minus one (Sánchez-Gutiérrez et al., 2018).

SUFF1_PFMF
SUFF2_PFMF
SUFF3_PFMF
SUFF4_PFMF

Percentage of other words in the family that are more frequent for the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4) suffix. It was computed by dividing the number of more frequent words in the family by the total number of members in the family minus one (Sánchez-Gutiérrez et al., 2018).

SUFF1_length
SUFF2_length
SUFF3_length
SUFF4_length

The number of letters of the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4) suffix (Sánchez-Gutiérrez et al., 2018).

SUFF1_P
SUFF2_P
SUFF3_P
SUFF4_P

Affix productivity measured by the probability that a given affix, i.e., the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4) will be encountered in a hapax (words that appear only once). It was computed by dividing all hapaxes in the corpus that contain a morpheme by the summed token frequency of a morpheme (Sánchez-Gutiérrez et al., 2018).

SUFF1_Px
SUFF2_Px
SUFF3_Px
SUFF4_Px

Affix productivity measured by the probability that a hapax (words that appear only once) contains a certain affix, i.e., the first (SUFF1), second (SUFF2), third (SUFF3), or fourth (SUFF4). It was computed by dividing all hapaxes in the corpus that contain a morpheme by the total of all hapax legomena in the corpus (Sánchez-Gutiérrez et al., 2018).

Response variables

LexicalD_RT_V_ELP
LexicalD_RT_V_ELP_z
LexicalD_ACC_V_ELP

The mean visual lexical decision latency (in ms) and its normalized (*z*-scored) version, and the proportion of accurate responses for a particular word across participants from the English Lexicon Project (Balota et al., 2007).

LexicalD_RT_V_ECP
LexicalD_RT_V_ECP_z
LexicalD_ACC_V_ECP

The mean latency (in ms) and its normalized version, and the proportion of accurate responses for a particular word in the word knowledge task across participants from the English Crowdsourcing Project (Mandera et al., 2020). This task is similar, but not identical, to the traditional lexical decision task. Participants were asked to indicate whether each item "is a word you know or not." Their results showed that RTs in this task correlate well with those from lexical decision in ELP and BLP, and hence we have labeled it as such. It should be noted that in this task, participants were not instructed to respond quickly, and were discouraged to guess (large penalty for labeling a nonword as a known word).

LexicalD_RT_V_BLP
LexicalD_RT_V_BLP_z
LexicalD_ACC_V_BLP

The mean visual lexical decision latency (in ms) and its normalized version, and the proportion of accurate responses for a particular word across participants from the British Lexicon Project (Keuleers et al., 2012).

LexicalD_RT_A_MALD
LexicalD_RT_A_MALD_z
LexicalD_ACC_A_MALD

The mean auditory lexical decision latency (in ms) and its normalized version, and the proportion of accurate responses for a particular word from the Massive Auditory Lexical Decision database (Tucker et al., 2019).

LexicalD_RT_A_AELP
LexicalD_RT_A_AELP_z
LexicalD_ACC_A_AELP

The mean auditory lexical decision latency (in ms) and its normalized version, and the proportion of accurate responses for a particular word from the Auditory English Lexicon Project (Goh et al., 2020).

Naming_RT_ELP
Naming_RT_ELP_z
Naming_ACC_ELP

The mean naming latency (in ms) and its normalized version, and the proportion of accurate responses for a particular word across participants from the English Lexicon Project (Balota et al., 2007).

SemanticD_RT_Calgary
SemanticD_RT_Calgary_z
SemanticD_ACC_Calgary

The mean latency (in ms) and its normalized version, and the proportion of accurate responses of concrete/abstract semantic decision (i.e., does the word refer to something concrete or abstract?) for a particular word from the Calgary database (Pexman et al., 2017).

Recog_Memory

Recognition memory performance indicated by d' (hits minus false alarms) (Khanna & Cortese, 2021).

Data analyses

As an initial effort, we examined the relationship between independent variables and reported the correlations between independent and dependent variables. We aimed to include the largest number of variables over the maximum number of words. Because variable values are available for partially different sets of words, including more variables leads to a smaller set of words, and selecting a larger set of words leads to a smaller variable set. As a compromise between these competing factors,

we created a subset of the data containing 1728 words with measurements on 130 independent variables (28 General, 17 Orthographic, 17 Phonological, 38 Semantic, 26 Orth-Phon, four Morphological) and 13 response variables (three visual lexical decision reaction times, three visual lexical decision accuracies, two auditory lexical decision times, two auditory lexical decision accuracies, one naming time, one naming accuracy, and one recognition memory). To create this subset, we excluded variables that are available for relatively small sets of words or those that have low overlap with other variables (in terms of the words that the values are available for; e.g., NFeatures, Feature_Perceptual, Emot_Assoc), are categorical (i.e., DPoS_Brys, DPoS_VanH, and First_Phon), or are in a vector form (e.g., Word2vec).

Interrelations between variables

Spearman's correlation between variables

To summarize the description of the relationship between each of the independent variables, we computed the Spearman's correlation among the 130 independent variables and created a similarity plot using these correlations.
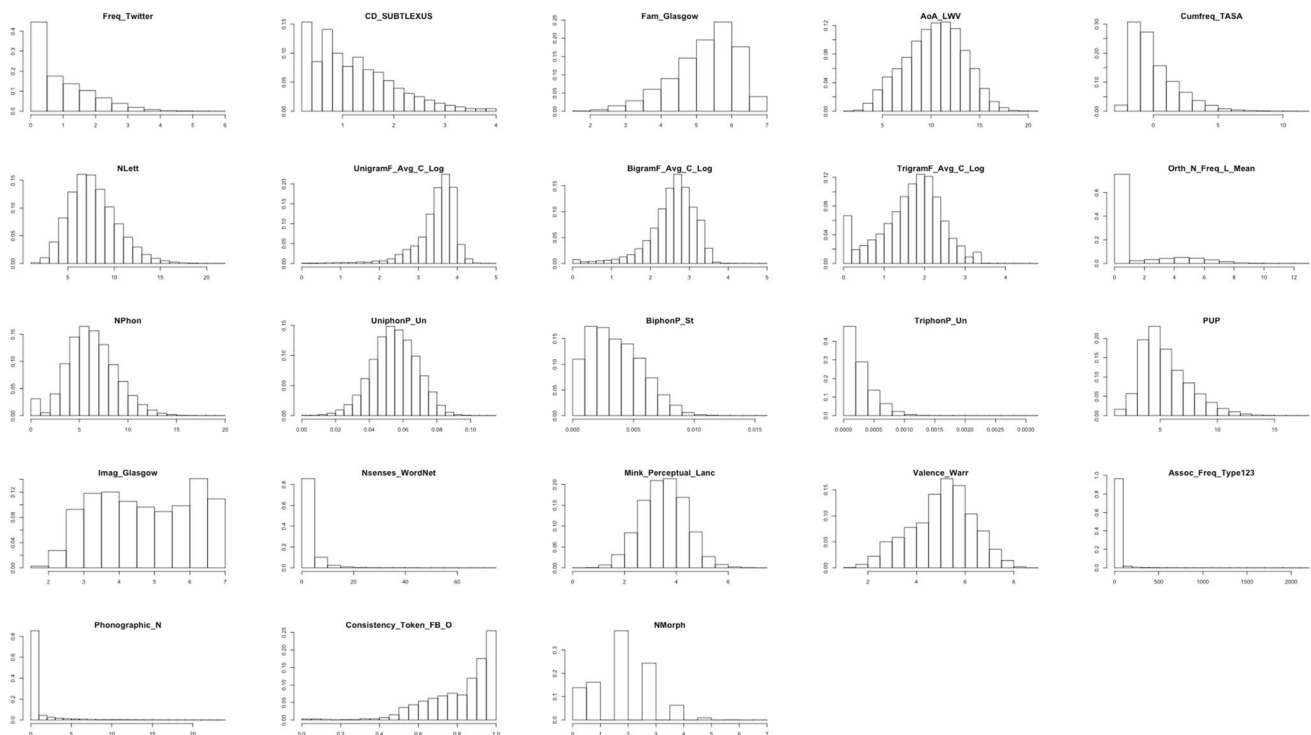

Fig. 1 Distribution of a sample of representative variables over all words for which the variable is available
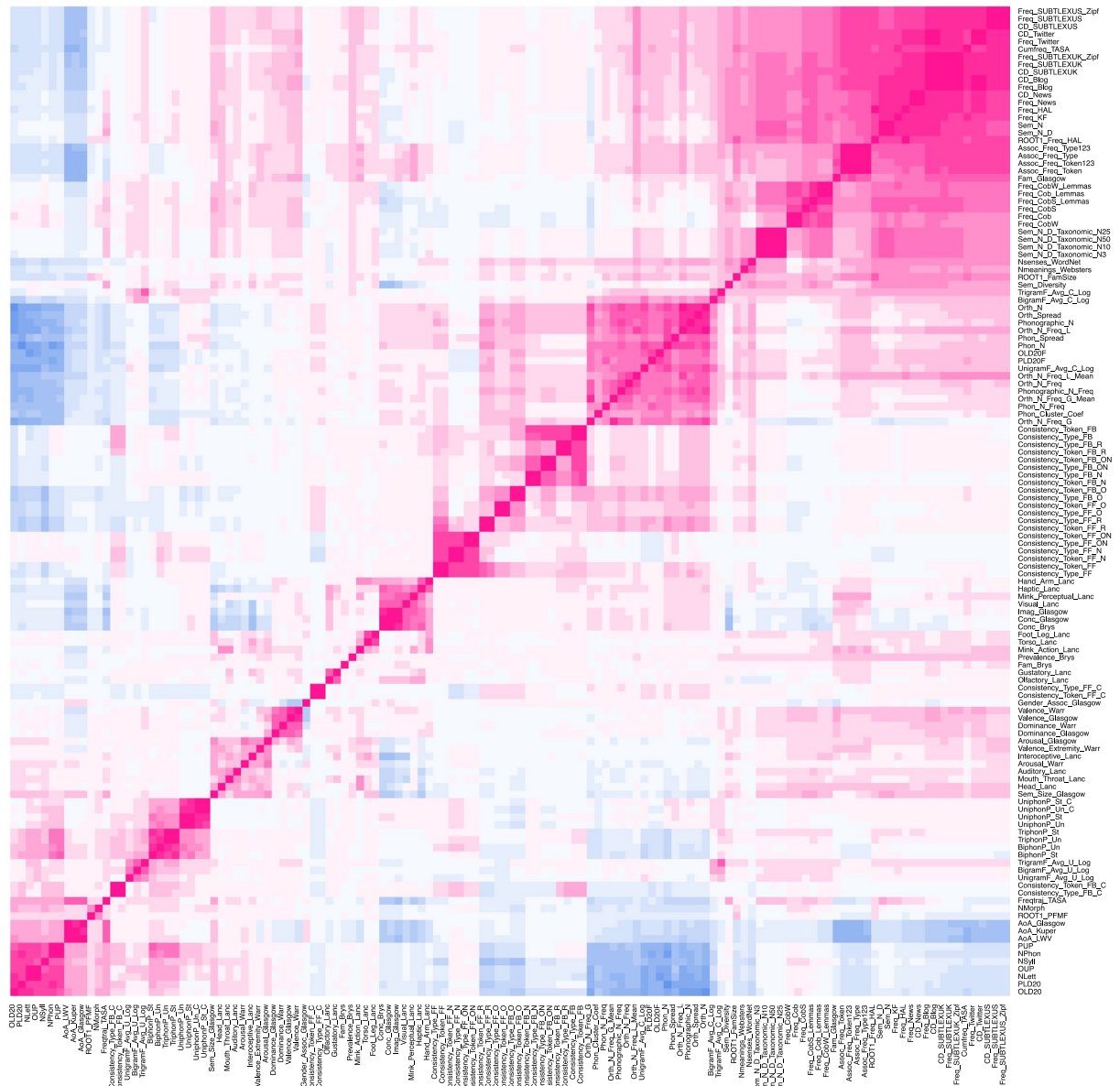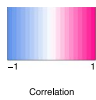
Fig. 2 Visualization of the similarity between 130 variables computed over 1728 words based on Spearman's correlation. Color represents correlation coefficients between different variables, ranging from –1 (*blue*) to 1 (*pink*)

t-Distributed Stochastic Neighbor Embedding (t-SNE)

We also examined the interrelations between variables with t-Distributed Stochastic Neighbor Embedding (t-SNE) and hierarchical cluster analyses. Barnes-Hut t-SNE (perplexity = 30; theta = 0) was used to visualize high-dimensional data. This method converts high-dimensional Euclidean distances between variables into conditional probabilities and then projects these distances onto a two-dimensional embedding space using the Student's *t* distribution by minimizing the Kullback–Leibler divergence. It has the advantage of revealing global structure while also capturing local structure of the high-dimensional data (Van der Maaten & Hinton, 2008).

Hierarchical cluster analysis

As an additional method of visualizing clustering among independent variables, hierarchical cluster analysis was performed using Ward's criterion (Murtagh & Legendre, 2014).

Exploratory factor analysis

A parallel analysis was performed to determine the appropriate number of latent factors (Crawford et al., 2010; Horn, 1965). The exploratory factor analysis was performed using the principal axis factoring extraction method and *oblimin* rotation.

Network analysis

To further determine the interrelations between variables, we performed a psychometric network modeling analysis in which each observed variable was modeled as a node and partial correlations among variables were modeled as edges (Epskamp et al.,

2018; Epskamp & Fried, 2018). Three centrality indices were computed to examine the relative importance of a node in the network: strength, closeness and betweenness. Strength refers to the sum of absolute partial correlation values for each node. Closeness refers to the inverse of the sum of distances from one node in the network to all other nodes. Betweenness refers to the number of the shortest paths that one node was passed through. A "best" measure of each individual variable group, defined as the measure that has the highest overall weighted correlation with dependent measures, was chosen to represent the group in the network analysis (e.g., CD_SUBTLEXUS for all contextual diversity measures). This ensures that that centrality indices are not biased towards one particular variable because of the unequal number of measures (e.g., having many frequency measures but few age-of-acquisition measures).

Correlations between variables/factors and the dependent variables

In addition to the analyses on the dataset that excluded the semantic decision task ($n = 1728$), we also performed
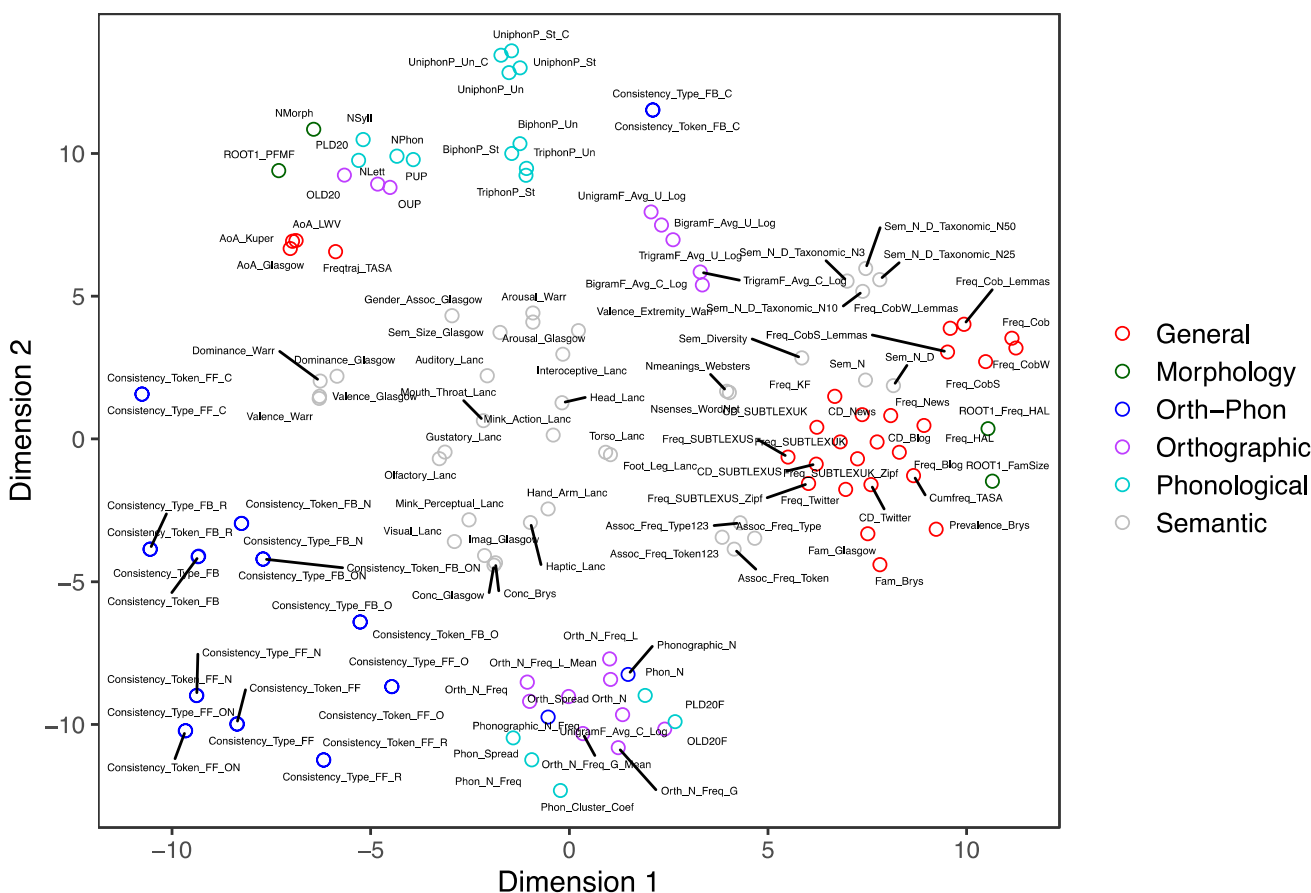

Fig. 3 Two-dimensional *t*-distributed stochastic neighbor embedding (t-SNE) results for 130 variables over 1728 words. *Colors* indicate six theoretically defined groups (i.e., General, Orthographic, Phonological, Semantic, Orth-Phon, Morphological)
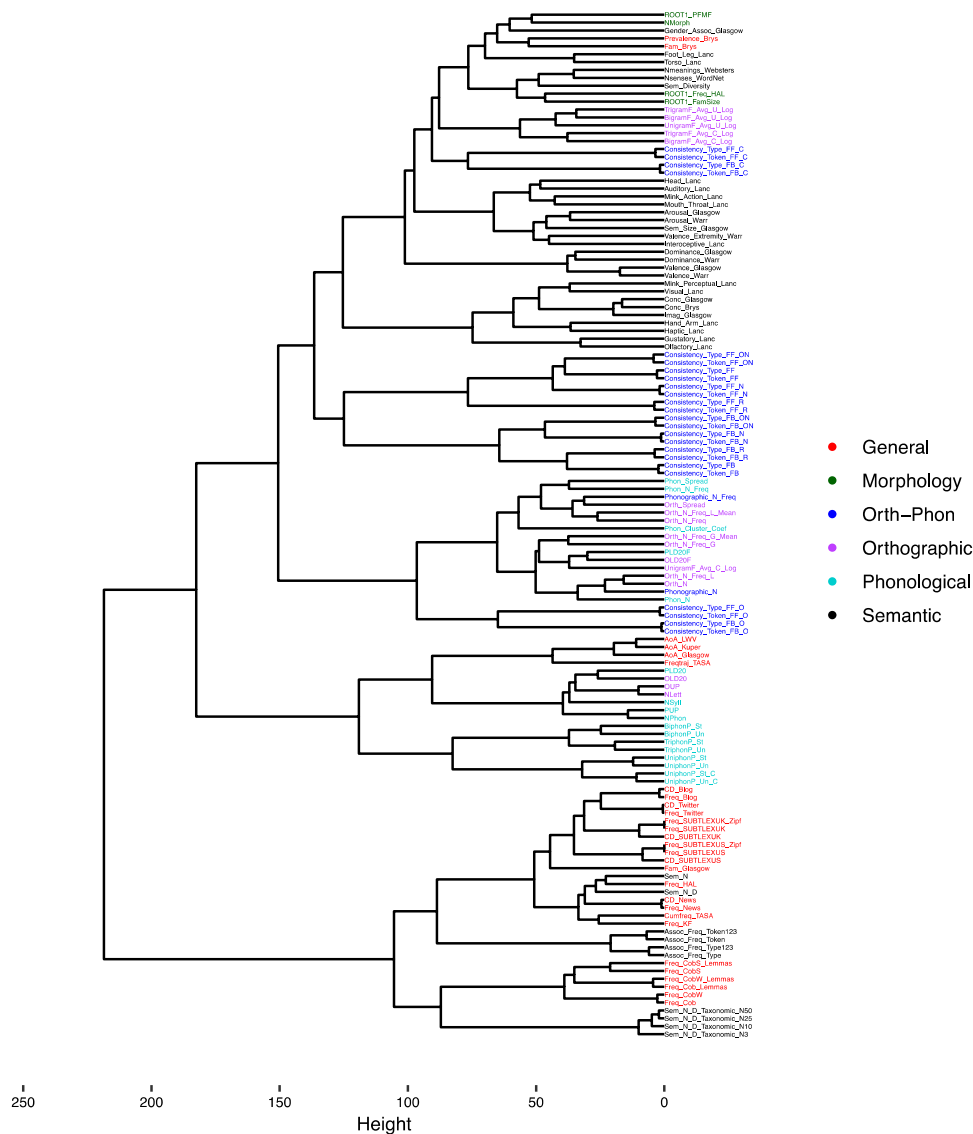
Fig. 4 Hierarchical clustering results using Ward's criterion for 130 variables over 1728 words

analyses for correlations between variables/factors and the dependent variables after including semantic decision (SemanticD_RT_Calgary), resulting in $n = 471$ words.

Correlations between each independent variable and the dependent variables

To examine bivariate relationships between the dependent and independent variables, normalized reaction time measures (zRTs) for each of seven dependent variables (i.e., LexicalD_RT_V_ELP_z, LexicalD_RT_V_ECP_z, LexicalD_RT_V_BLP_z, LexicalD_RT_A_MALD_z, LexicalD_RT_A_AELP_z, Naming_RT_ELP_z, and Recog_Memory) were correlated with each of the 130 variables over 1728 words using Spearman's correlation. In addition

to correlations between independent variables and each of the dependent variables separately, we also computed an overall weighted absolute correlation that gives equal weight to each task (visual lexical decision, auditory lexical decision, naming, and recognition memory), so that the overall value is not dominated by tasks such as visual lexical decision that have multiple measures. It was computed using Spearman's R for each measure as [(visual lexical decision R of ELP + visual lexical decision R of ECP + visual lexical decision R of BLP)/3 + (auditory lexical decision R of MALD + auditory lexical decision R of AELP)/2 + naming R + recognition memory R]/4. A similar weighted absolute correlation was computed for the smaller dataset ($n = 471$) that included semantic decision times, using five different tasks. We also provide ranks of different measures for each

Table 1 Variables with factor loadings larger than 0.4 for the exploratory factor analysis over 1728 words

PA1: Freq_CD	PA3: Consistency_FF_ONN	PA7: Sem_N_Taxonomic	PA15: Assoc_Freq	PA13: Freq_Cob	PA6: Consistency_FB_ONN
Freq_HAL	Consistency_Token_FF_ON	Sem_N_D_Taxonomic_N10	Assoc_Freq_Token123	Freq_Cob	Consistency_Type_FB
Freq_KF	Consistency_Type_FF_ON	Sem_N_D_Taxonomic_N25	Assoc_Freq_Token	Freq_CobW	Consistency_Token_FB
Freq_SUBTLEXUS	Consistency_Token_FF_N	Sem_N_D_Taxonomic_N50	Assoc_Freq_Type	Freq_Cob_Lemmas	Consistency_Token_FB_ON
Freq_SUBTLEXUS_Zipf	Consistency_Type_FF_N	Sem_N_D_Taxonomic_N3	Assoc_Freq_Type123	Freq_CobS	Consistency_Token_FB_N
Freq_SUBTLEXUK	Consistency_Token_FF			Freq_CobW_Lemmas	Consistency_Type_FB_N
Freq_SUBTLEXUK_Zipf	Consistency_Type_FF			Freq_CobS_Lemmas	Consistency_Type_FB_ON
Freq_Blog					
Freq_Twitter					
Freq_News					
CD_SUBTLEXUS					
CD_SUBTLEXUK					
CD_Blog					
CD_Twitter					
CD_News					
Fam_Glasgow					
Cumfreq_TASA					
Sem_N					
Sem_N_D					
PA4: ConcImage	**PA22: OrthPhon_OLD**	**PA5: Uniphon_P**	**PA23: OrthPhon_N**	**PA17: BiTriPhon_P**	**PA8: Valence_Dominance**
Interoceptive_Lanc	OLD20	UniphonP_Un_C	Orth_N	TriphonP_Un	Valence_Glasgow
Sem_Diversity	Orth_N_Freq	UniphonP_St_C	Phonographic_N	TriphonP_St	Valence_Warr
Conc_Glasgow	Phonographic_N_Freq	UniphonP_St	Orth_N_Freq_L	BiphonP_St	Dominance_Warr
Imag_Glasgow	Orth_N_Freq_L_Mean	UniphonP_Un	Phon_N	BiphonP_Un	Dominance_Glasgow
Conc_Brys	Orth_Spread		Orth_N_Freq_G		
Visual_Lanc	Phon_Spread				
Mink_Perceptual_Lanc	Phon_N_Freq				
	Orth_N_Freq_G_Mean				
PA9: Consistancy_FB_CR	**PA20: AoA**	**PA2: NLettPhon_Unique**	**PA19: Consistency_FF_R**	**PA12: NGramLog**	**PA21: Consistency_FF_O**
Consistency_Type_FB_R	AoA_LWV	OLD20F	Consistency_Type_FF_R	BigramF_Avg_U_Log	Consistency_Type_FF_O
Consistency_Token_FB_R	AoA_Glasgow	PLD20F	Consistency_Token_FF_R	TrigramF_Avg_U_Log	Consistency_Token_FF_O
Consistency_Token_FB_C	AoA_Kuper	NLett		UnigramF_Avg_U_Log	
Consistency_Type_FB_C	Freqtraj_TASA	OUP		BigramF_Avg_C_Log	
Consistency_Type_FB		NPhon		TrigramF_Avg_C_Log	
Consistency_Token_FB		PUP			
PA11: Action	**PA18: Arousal**	**PA10: Consistency_FB_O**	**PA14: Consistency_FF_C**	**PA16: Gust_Olfac**	**PA24: Haptic**
Hand_Arm_Lanc	Arousal_Glasgow	Consistency_Token_FB_O	Consistency_Token_FF_C	Gustatory_Lanc	Head_Lanc
Mink_Action_Lanc	Arousal_Warr	Consistency_Type_FB_O	Consistency_Type_FF_C	Olfactory_Lanc	Auditory_Lanc
Torso_Lanc	Sem_Size_Glasgow			Mouth_Throat_Lanc	Haptic_Lanc
Foot_Leg_Lanc	Valence_Extremity_Warr				Hand_Arm_Lanc

variable (e.g., frequency) by overall weighted correlation or correlation values for response variables from each dataset (e.g., ELP, ECP etc.). The measure of each independent variable that has the highest overall weighted correlation was chosen for the network analysis.

Correlations between each factor and the dependent variables

We also performed Spearman's correlation between factor scores of each factor obtained from the exploratory factor
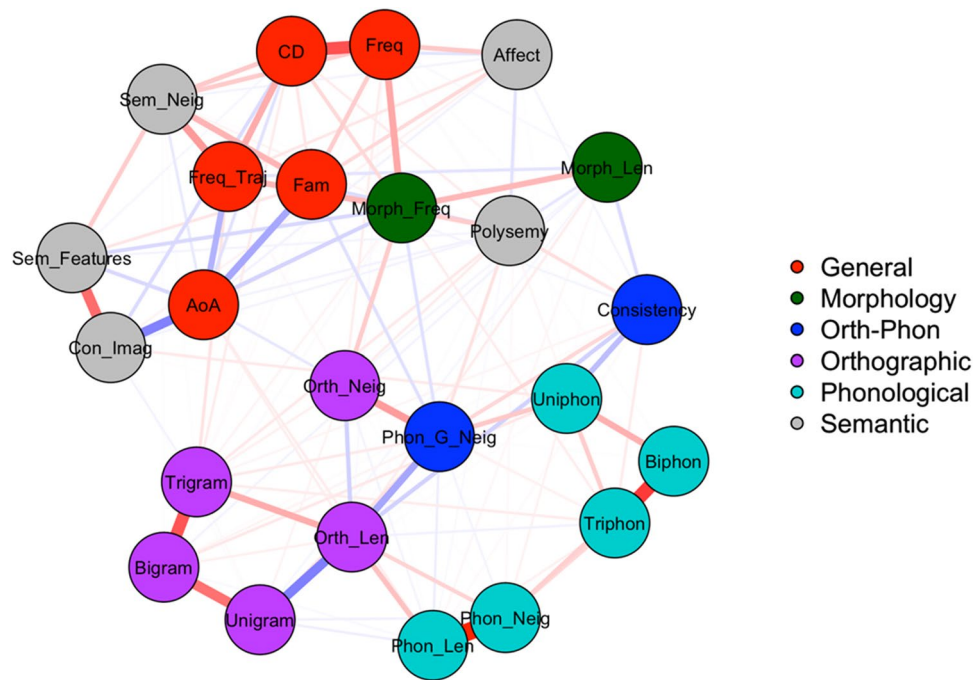
Fig. 5 Psychometric network model describing relationships between independent variables over 1728 words. *Line thickness* refers to the edge strength, which is the size of the partial correlation between var- iables. *Line colors* indicate the sign of the correlation, in which *red lines* correspond to a positive correlation while *blue lines* correspond to a negative correlation

analysis, and normalized reaction time measures (zRTs) for each of seven dependent variables (i.e., LexicalD_RT_V_ELP_z, LexicalD_RT_V_ECP_z, LexicalD_RT_V_BLP_z, LexicalD_RT_A_MALD_z, LexicalD_RT_A_AELP_z, Naming_RT_ELP_z, and Recog_Memory).

Contributions of distributional semantic vectors to the dependent variables

We also performed multiple regression analyses with three distribution semantic vectors (Word2Vec, GloVe, and Taxonomic) as predictors and each of the seven dependent variables as response variables. Adjusted multiple R was obtained from each of the multiple regression analyses for comparison with correlation values of other variables. Overall weighted R for the distributional semantic vectors was computed in a similar way as for other independent variables.

Correlations between each independent variable and dependent variables for nonwords

To further examine the bivariate relationships between the dependent and independent variables, we also correlated normalized reaction time measures (zRTs) of *nonwords* for LexicalD_RT_V_ELP_z, LexicalD_RT_V_BLP_z, Lexi- calD_RT_A_MALD_z with a set of independent variables.

These variables include NLett, Orth_N, OLD20, OUP, Orth_ Spread obtained using LexiCAL (Chee et al., 2021); and UnigramF_Avg_C_Log, BigramF_Avg_C_Log, TrigramF_ Avg_C_Log, UnigramF_Avg_U_Log, BigramF_Avg_U_Log, and TrigramF_Avg_U_Log retrieved from MCWord database (Medler & Binder, 2005). Given that IPA transcription is avail- able for AELP database (Goh et al., 2020), we were able to compute additional measures including NPhon, NSyll, PLD20, PUP, Phon_Spread, and Phonographic_N using LexiCAL for AELP. We then correlated these measures with reaction time measure of nonwords for LexicalD_RT_A_AELP.

To compare the correlations between reaction time meas- ures and independent variables across databases, we merged across databases and compared the databases that have a rea- sonable sample size of overlapping nonwords. This resulted in 1292 nonwords shared between ELP and BLP, 480 non- words shared between ELP and AELP, and 574 nonwords shared between AELP and BLP.

Results

We present the distribution for a sample of representative variables in Fig. 1. The distribution of all 143 variables for all available words for that variable is shown separately for each of the seven groups (i.e., General, Orthographic, Pho- nological, Semantic, Orth-Phon, Morphological, Response)
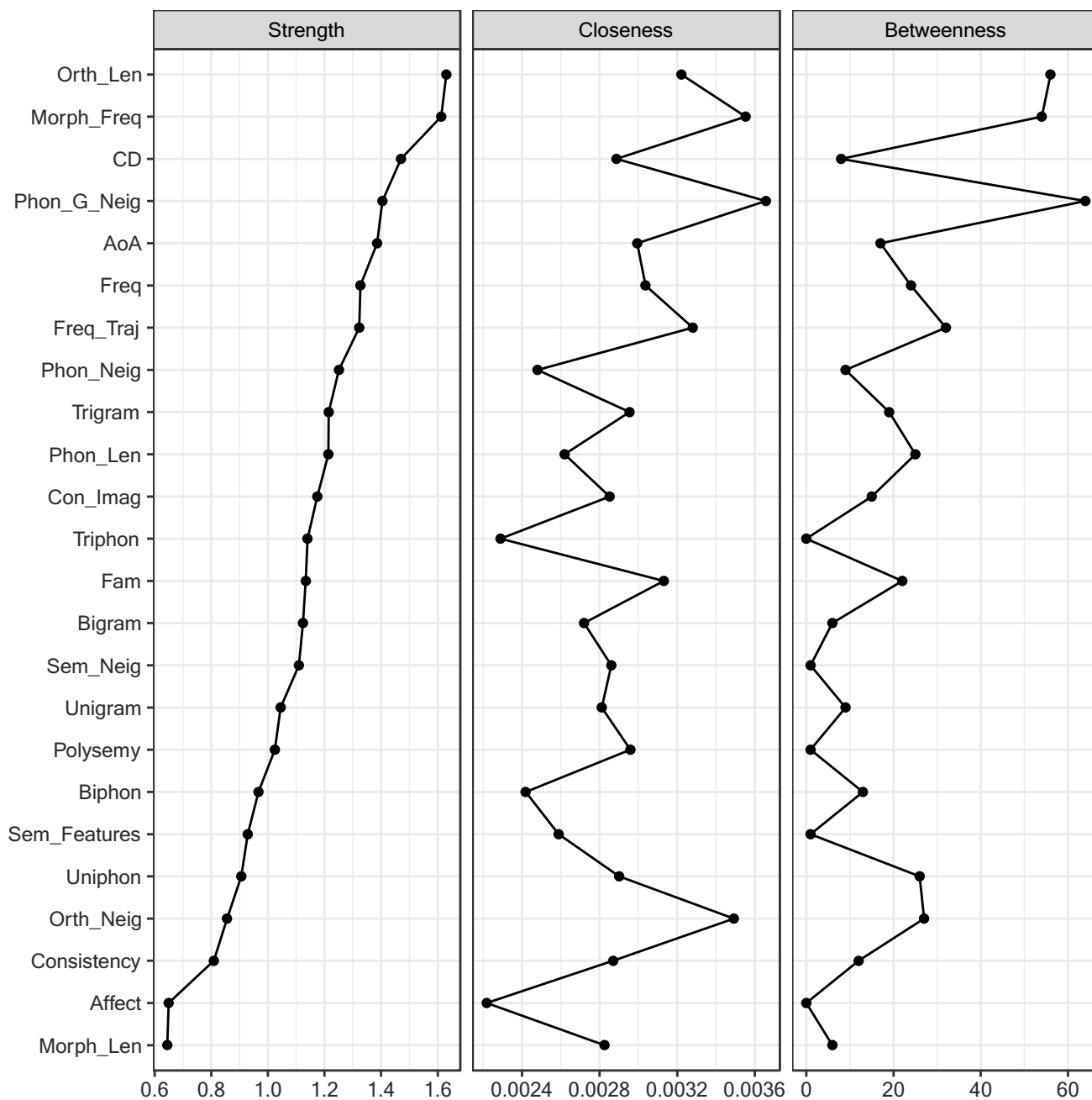
Fig. 6 Three centrality indices of the psychometric network model over 1728 words: strength, closeness and betweenness. Strength refers to the sum of absolute partial correlation values for each node. Closeness refers to the inverse of the sum of distances from one node in the network to all other nodes. Betweenness refers to the number of the shortest paths that one node was passed through. The variables are ordered in a descending order by strength

in Supplemental Figures 1–7 (Supplemental material can be found at https://osf.io/9qbjz/). Frequency measures from SUBTLEX and Worldlex were widely distributed across the whole range. Compared to constrained unigram, bigram, or trigram frequencies, values for the unconstrained versions of frequency measures were more distributed. The concreteness measures had relatively uniform distributions. The measures of reaction times generally had skewed distributions as expected. The accuracies were generally very high.

Interrelations between variables

Spearman's correlations between variables

The similarity among 130 predictor variables based on Spearman's correlation is shown in Fig. 2 Several clusters can be identified from this visualization. The largest cluster included a series of frequency measures such as those from SUBTLEX, CELEX, HAL and Worldlex databases;
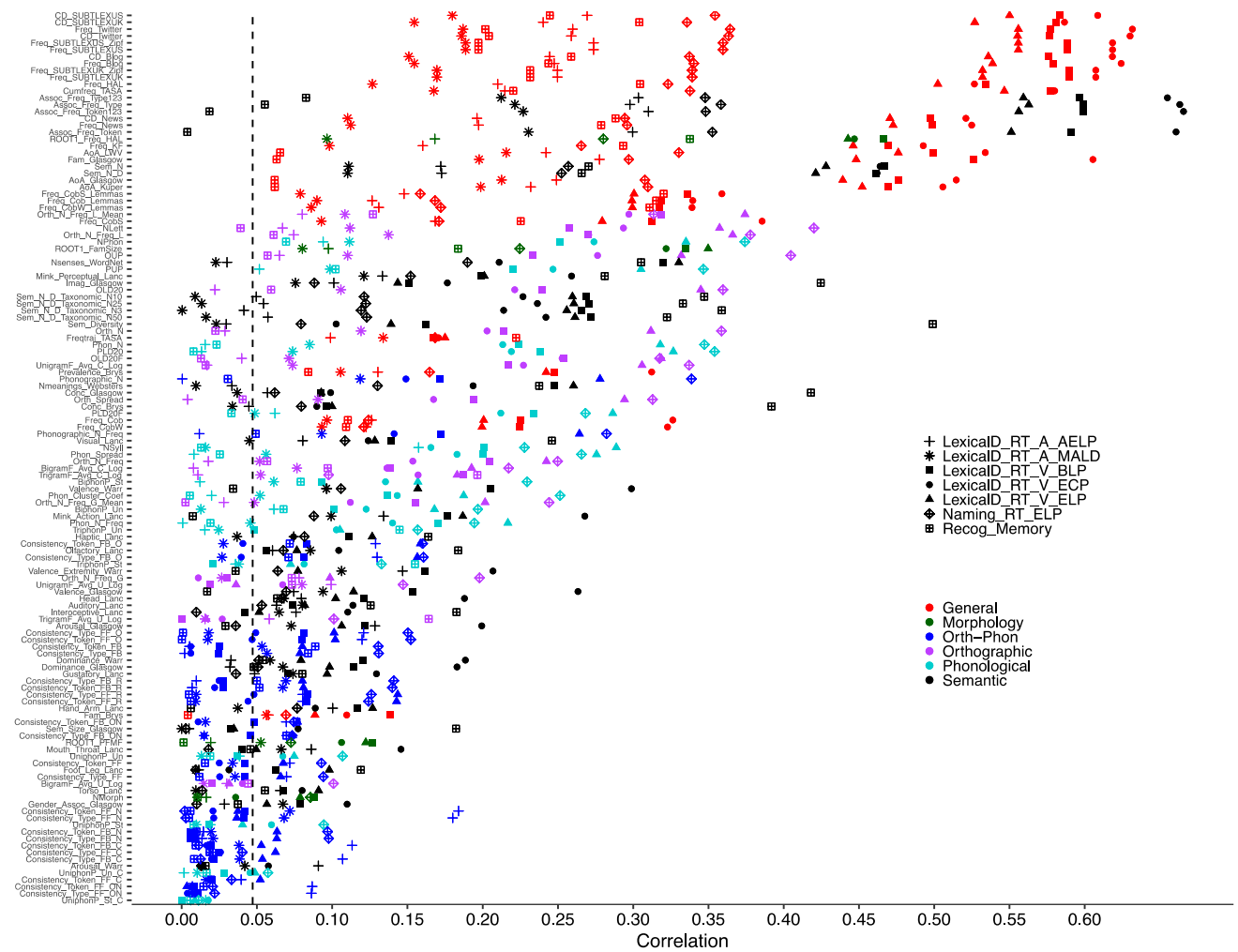
Fig. 7 Absolute Spearman's correlation values for 130 variables with each of the seven dependent variables (six zRTs and d' for recognition memory) over 1728 words. The variables are ordered in a descending order by overall weighted absolute correlation values. The *vertical dashed line* indicates a threshold for $p < 0.05$ level of significance

cumulative frequency; semantic neighborhood measures; association frequency measures and familiarity. The second cluster included orthographic neighborhood measures, phonological neighborhood measures, frequencies for the orthographic and phonological neighbors, orthographic, phonological and phonographic neighborhood measures, as well as orthographic and phonological spread measures. The third cluster included length and uniqueness-related variables, such as the number of letters, number of syllables, number of phonemes, number of morphemes, orthographic and phonological Levenshtein distances, orthographic and phonological uniqueness point, and age of acquisition measures. The fourth cluster included a series of positional probability measures, biphon and triphon probability measures. The fifth cluster included sensory and motor semantic variables, including visual features, haptic features, Minkowski perceptual strength, imageability, concreteness, and strength of

experiences with hand/arm. The last cluster as also semantic, and included strength of experiences with head or mouth/throat, auditory feature, interoceptive feature, arousal, and semantic size (Fig. 2). Similarity plots for each of the General, Orthographic, Phonological and Semantic groups are also shown in Supplemental Material (Supplemental Figures 8–11; other groups are not shown due to small number of variables).

t-Distributed Stochastic Neighbor Embedding (t-SNE)

Variable groupings based on t-SNE visualization partially followed our categorical assignments (Fig. 3). Most general variables clustered together, with those related to development (age of acquisition and frequency trajectory) forming a distinct cluster. Orthographic variables also clustered together, with those related to unconstrained and constrained

Table 2 Absolute correlation values of the 'best' independent variable within each subcategory with seven dependent variables as well as rankings over 1728 words

Variable name	Group (category)	Group (subcategory)	Weighted_ Overall_R	Weighted_ Overall_ Rank
Freq_Twitter	General	Frequency	0.344	6
CD_SUBTLEXUS	General	Contextual Diversity	0.351	4
Fam_Glasgow	General	Familiarity	0.281	25
AoA_LWV	General	Age of Acquisition	0.282	24
Cumfreq_TASA	General	Frequency Trajectory	0.329	15
NLett	Orthographic	Orthographic Length	0.213	35
UnigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Unigram	0.162	55
BigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Bigram	0.128	70
TrigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Trigram	0.125	71
Orth_N_Freq_L_Mean	Orthographic	Orthographic Neighborhood	0.216	33
NPhon	Phonological	Phonological Length	0.208	37
UniphonP_Un	Phonological	Phonotactic Probabilities: Uniphon	0.053	112
BiphonP_St	Phonological	Phonotactic Probabilities: Biphon	0.118	72
TriphonP_Un	Phonological	Phonotactic Probabilities: Triphon	0.102	79
PUP	Phonological	Phonological Neighborhood	0.195	41
Imag_Glasgow	Semantic	Concreteness/Imageability	0.189	43
Nsenses_WordNet	Semantic	Polysemy	0.202	40
Mink_Perceptual_Lanc	Semantic	Specific Semantic Features	0.195	42
Valence_Warr	Semantic	Affect	0.114	73
Assoc_Freq_Type123	Semantic	Semantic Neighborhood	0.323	16
Phonographic_N	Orth-Phon	Phonographic Neighborhood	0.157	57
Consistency_Token_FB_O	Orth-Phon	Consistency	0.101	81
NMorph	Morphology	Morphological Length	0.044	118
ROOT1_Freq_HAL	Morphology	Morphological Frequency	0.301	22
GloVe	Semantic	Vector Representation	0.484	1

Weighted_Overall_R indicates the overall weighted absolute correlation that gives equal weight to each task (visual lexical decision, auditory lexical decision, naming, and recognition memory). Weighted_Overall_Rank indicates the rank based on Weighted_Overall_R among all 130 variables

ngrams forming a different cluster. Phonological variables formed two categories, with one related to neighborhood measures, while the other mainly related to phonotactic variables and length. Semantic variables were spread out and formed three clusters, with the largest cluster, with sensory-motor and affective features, occupying the center of the space. A second cluster was related to semantic neighborhood measures, while the third was related to the frequency of the associates. Though Orth-Phon variables were mostly related to consistency, they were still relatively spread out, suggesting that different aspects of consistency capture different properties, and should not be treated in a unitary manner.

Hierarchical cluster analysis

Hierarchical clustering showed similar groupings as the similarity and t-SNE results (Fig. 4). This visualization is useful for identifying within- and between-category similarities.

For example, some semantic variables clustered with General variables, while others with Phonological and Orthographic variables.

Exploratory factor analysis

The exploratory factor analysis with a number of 24 latent factors determined by the parallel analysis was performed. Variables with factor loadings larger than 0.4 are presented in Table 1. A full table with factor loading values is provided in Supplemental Table 2. Based on the table of factor loadings, we labeled these factors as following: Freq_CD, Consistency_FF_ONN, Sem_N_Taxonomic, Assoc_Freq, Freq_Cob, Consistency_FB_ONN, ConcImage, OrthPhon_OLD, Uniphon_P, OrthPhon_N, BiTriPhon_P, Valence_Dominance, Consistency_FB_CR, AoA, NLettPhon_Unique, Consistency_FF_R, NGramLog, Consistency_FF_O, Action, Arousal, Consistency_FB_O, Consistency_FF_C, Gust_Olfac, Haptic.

Table 3 Absolute correlation values of the 'best' independent variable within each subcategory with eight dependent variables including Semantic Decision, as well as rankings, over 471 words

Variable name	Group (category)	Group (subcategory)	Weighted_Overall_R	Weighted_Overall_Rank
Freq_SUBTLEXUS	General	Frequency	0.357	5
CD_SUBTLEXUS	General	Contextual Diversity	0.356	7
Fam_Glasgow	General	Familiarity	0.339	21
AoA_LWV	General	Age of Acquisition	0.361	4
Cumfreq_TASA	General	Frequency Trajectory	0.352	10
NLett	Orthographic	Orthographic Length	0.168	49
UnigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Unigram	0.146	61
BigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Bigram	0.143	64
TrigramF_Avg_C_Log	Orthographic	Graphotactic Probabilities: Trigram	0.134	68
Orth_N_Freq_L	Orthographic	Orthographic Neighborhood	0.181	47
NPhon	Phonological	Phonological Length	0.166	51
UniphonP_Un	Phonological	Phonotactic Probabilities: Uniphon	0.058	111
BiphonP_St	Phonological	Phonotactic Probabilities: Biphon	0.082	96
TriphonP_Un	Phonological	Phonotactic Probabilities: Triphon	0.063	109
PLD20	Phonological	Phonological Neighborhood	0.160	53
Imag_Glasgow	Semantic	Concreteness/Imageability	0.266	28
Nsenses_WordNet	Semantic	Polysemy	0.155	54
Mink_Perceptual_Lanc	Semantic	Specific Semantic Features	0.275	26
Valence_Warr	Semantic	Affect	0.133	70
Assoc_Freq_Token123	Semantic	Semantic Neighborhood	0.349	11
Phonographic_N	Orth-Phon	Phonographic Neighborhood	0.127	74
Consistency_Token_FF_R	Orth-Phon	Consistency	0.114	80
NMorph	Morphology	Morphological Length	0.019	133
ROOT1_Freq_HAL	Morphology	Morphological Frequency	0.272	27
Word2Vec	Semantic	Vector Representation	0.567	1

Weighted_Overall_R indicates the overall weighted absolute correlation that gives equal weight to each task (visual lexical decision, auditory lexical decision, naming, recognition memory, and semantic decision). Weighted_Overall_Rank indicates the rank based on Weighted_Overall_R among all 130 variables

Network analysis

The network analysis showed that the clusters partially reflected the theoretically defined groups: General, Orthographic, Phonological, Semantic, Orth-Phon, and Morphological (Fig. 5). Semantic features were especially distributed. While sensory-motor semantic features and concreteness cluster together, other semantic groups representing semantic neighborhood, affect, and polysemy were distinct from them as well as from each other. Overall, morphological frequency and orthographic length were the variables that most strongly connected to other variables (Fig. 6). Notably, phonographic neighborhood was the variable most close to other nodes in the network and a hub that other variables passed through. Contextual diversity, frequency and age of acquisition were also among the variables that had strong connections with other variables. We also performed network analyses on the data after including semantic decision with *n* = 471 words.

The psychometric network model and centrality indices for these analyses are provided in Supplemental Figures 12 and 13.

Correlations between variables/factors and the dependent variables

Correlations between each independent variable and the dependent variables

The Spearman's correlation between 130 variables and each of the seven dependent variables of reaction times is shown in Fig. 7. Separate correlation plots for each dependent measure are presented in supplemental material (Supplemental Figures 14–20). The absolute correlation values of the 'best' variable (highest correlation) in each group, and its ranking, is given in Table 2. The full table containing correlations and rankings for each task for each variable is provided in Supplemental Table 3.
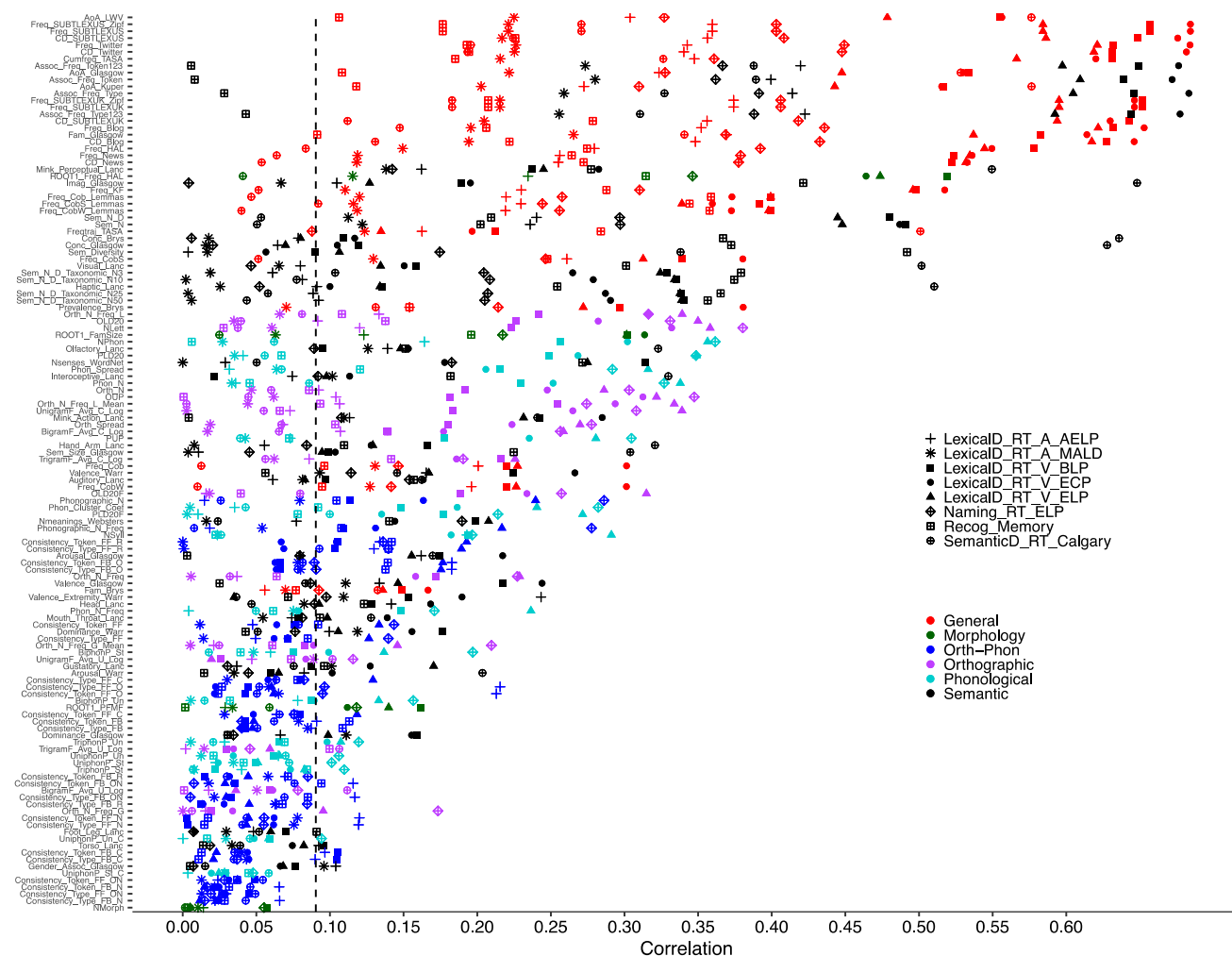
Fig. 8 Absolute Spearman's correlation values for 130 variables with each of the eight dependent variables (seven zRTs and d' for recognition memory) over 471 words. The variables are ordered in a descending order by overall weighted absolute correlation values. The *vertical dashed line* indicates a threshold for $p < 0.05$ level of significance

Overall, General variables had the largest correlation with visual and auditory lexical decision reaction times. Contextual diversity, along with frequency measures, were overall the most significant predictors. Association frequency measures were also among the most significant predictors, especially for visual and auditory lexical decision tasks. Overall, auditory lexical decision correlations were much lower than those of other measures. Age of acquisition and familiarity were near the top for auditory lexical decision, differentiating this task from other measures.

A subgroup of Semantic variables was the next most informative overall, followed by several Orthographic and a few Phonological variables. Orth-Phon variables, as a group, were the least informative on average. However, we note that more emphasis was given to lexical decision tasks (due to the presence of both visual and auditory lexical decision), whereas Orth-Phon variables were relevant especially for

naming. Indeed, they showed much higher correlation in the naming task, but still lagged behind several General, Semantic, and Orthographic variables. Naming was the only task where Orthographic variables (length and uniqueness point) were strongest predictors (Supplementary Figure 17). Recognition memory was differentiated by the fact that semantic diversity and imageability/concreteness were the top predictors, followed by taxonomic semantic neighborhood measures (Supplementary Figure 20). Unlike other measures, lemma frequencies were found to be more predictive than wordform frequencies for recognition memory.

For the 471 words when semantic decision was included, the results for lexical decision reaction times and naming latencies were generally consistent with previous findings with the larger subset of words shown in Fig. 7. The correlation values of the 'best' variable in each group, and its ranking, is given in Table 3, with the full table provided
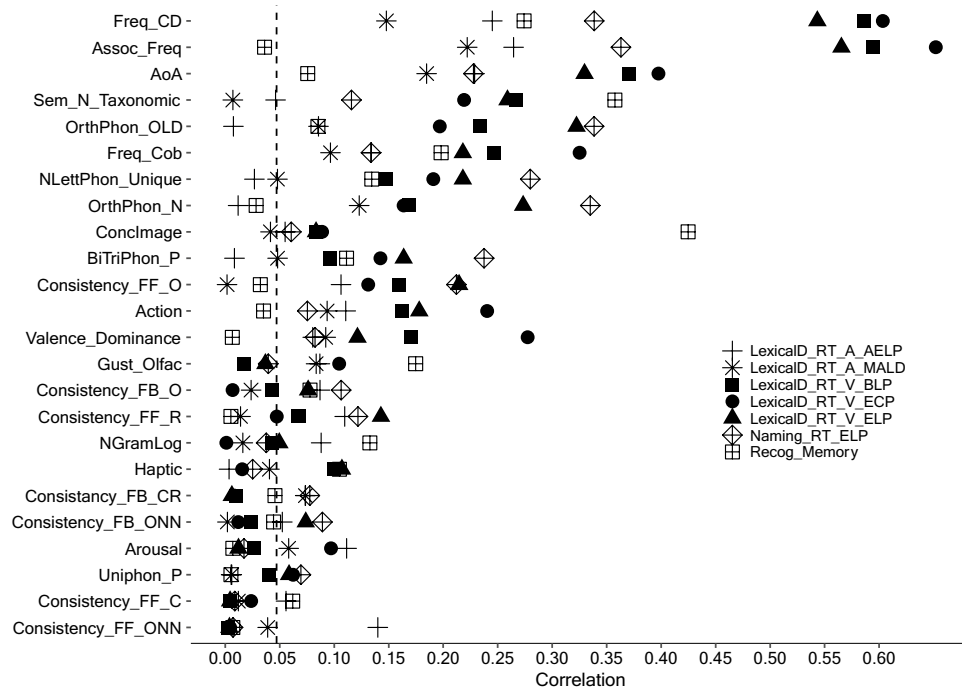
Fig. 9 Absolute correlation values for 24 factors with each of the seven dependent variables (six zRTs and d' for recognition memory) over 1728 words. The variables are ordered in a descending order by overall weighted absolute correlation values. The *vertical dashed line* indicates a threshold for $p < 0.05$ level of significance

in Supplemental Table 4. The overall ranking of the variables changes somewhat due to the inclusion of the semantic decision task. The strongest predictors for semantic decision were Semantic and General variables such as concreteness/imageability, and age of acquisition measures. In contrast to other measures, frequency and contextual diversity were ranked relatively lower for semantic decision (Fig. 8). Correlation plot for semantic decision reaction times over 471 words was presented in Supplemental Figure 21.

Correlations between each factor and the dependent variables

The correlations between 24 factors and each of the seven dependent variables of reaction times are shown in Fig. 9 and Table 4. Same as the correlation between 130 individual variables and the seven dependent variables, factors representing general variables such as frequency and contextual diversity had the largest correlation with visual and auditory lexical decision reaction times. Orth-Phon variables were also relevant especially for naming. We also found a high contribution of imageability/concreteness and semantic neighborhood to recognition memory.

To examine the correlations between each factor and the dependent variables for the 471 words when semantic decision was included, we first ran an exploratory factor analysis with a number of 19 latent factors determined

by the parallel analysis. Factor loadings (> 0.4) are presented in Supplemental Table 5. Based on the table of factor loadings, we named these factors as follows: Freq_CD, OrthPhon_N, ConcImage_AoA, Consistency_FB_ONNR, Consistency_FF_ONNR,

Sem_N_Taxonomic, Freq_Cob, Assoc_Freq, Uniphon_P, Valence_Dominance, BiTriPhon_P, Consistency_FB_C, Consistency_FB_O, Consistency_FF_C, BiTrigramLog, Gust_Olfac, Action, Consistency_FF_O, and LowArousal. The correlations between 19 factors and each of the 7 dependent variables of reaction times are shown in Fig. 10 and Table 5. Similarly, we found that General variables such as frequency and contextual diversity had the largest correlation with visual and auditory lexical decision reaction times and semantic neighborhood was highly correlated with recognition memory. For semantic decision, the strongest factor was concreteness/imageability, which was combined with other AoA.

Contributions of distributional semantic vectors to the dependent variables

The multiple regression analyses with three distribution semantic vectors (Word2Vec, GloVe, and Taxonomic) as predictors and each of the seven dependent variables as response variables showed that the overall weighted values of adjusted multiple R were 0.481 for Word2Vec, 0.484 for GloVe, and 0.462 for Taxonomic. These semantic

Table 4 Actual correlation values for 24 factors with each of the seven dependent variables (six zRTs and d' for recognition memory) over 1728 words

Factor_ID	FactorNames	Weighted_ Overall_R	LexicalD_ RT_V_ ELP_R	LexicalD_ RT_V_ ECP_R	LexicalD_ RT_V_ BLP_R	LexicalD_RT_A_ MALD_R	LexicalD_RT_A_ AELP_R	Naming_RT_ ELP_R	Recog_ Memory_R
PA1	Freq_CD	–0.347	–0.543	–0.603	–0.586	–0.148	–0.245	–0.339	–0.274
PA15	Assoc_Freq	–0.312	–0.565	–0.652	–0.594	–0.222	–0.265	–0.363	–0.036
PA20	AoA	0.219	0.33	0.398	0.371	0.185	0.229	0.228	–0.076
PA7	Sem_N_Taxo-nomic	–0.187	–0.259	–0.219	–0.267	–0.007	–0.046	–0.116	–0.358
PA22	OrthPhon_ OLD	–0.18	–0.322	–0.197	–0.234	–0.086	0.008	–0.338	–0.085
PA13	Freq_Cob	–0.178	–0.218	–0.325	–0.247	–0.097	–0.134	–0.134	–0.198
PA2	NLettPhon_ Unique	0.159	0.218	0.191	0.148	0.048	0.027	0.28	–0.134
PA23	OrthPhon_N	–0.158	–0.273	–0.164	–0.169	–0.123	–0.012	–0.335	–0.028
PA4	ConcImage	–0.155	–0.083	–0.089	–0.083	–0.041	–0.055	–0.061	0.425
PA17	BiTriPhon_P	0.128	0.164	0.142	0.096	0.048	–0.008	0.238	–0.111
PA21	Consistency_ FF_O	–0.117	–0.215	–0.131	–0.16	–0.002	0.106	–0.212	–0.032
PA11	Action	–0.102	–0.178	–0.24	–0.163	–0.094	–0.111	–0.075	–0.035
PA8	Valence_Dom-inance	–0.091	–0.121	–0.278	–0.17	–0.092	–0.08	–0.082	0.007
PA16	Gust_Olfac	–0.088	–0.037	–0.105	–0.017	–0.083	–0.087	–0.039	0.175
PA10	Consistency_ FB_O	–0.07	–0.076	0.007	–0.043	–0.024	0.087	–0.106	–0.078
PA19	Consistency_ FF_R	–0.069	–0.143	–0.047	–0.068	–0.014	0.11	–0.122	0.005
PA12	NGramLog	–0.063	–0.05	0.001	–0.043	0.016	0.088	0.038	–0.133
PA24	Haptic	–0.057	–0.107	0.016	–0.099	0.041	0.004	–0.025	–0.105
PA9	Consistancy_ FB_CR	–0.051	–0.006	0.008	0.01	–0.074	–0.073	–0.078	–0.046
PA6	Consistency_ FB_ONN	–0.049	–0.074	0.012	–0.024	–0.002	0.052	–0.089	–0.044
PA18	Arousal	0.039	–0.012	–0.097	–0.026	–0.058	–0.111	0.017	0.007
PA5	Uniphon_P	0.033	0.059	0.062	0.04	–0.006	–0.005	0.069	0.005
PA14	Consistency_ FF_C	–0.029	–0.004	0.024	–0.004	–0.012	–0.056	0.009	–0.062
PA3	Consistency_ FF_ONN	–0.027	0.004	–0.004	–0.003	–0.039	–0.14	–0.007	–0.007

Weighted_Overall_R indicates the overall weighted correlation that gives equal weight to each task (visual lexical decision, auditory lexical decision, naming, and recognition memory). LexicalD_RT_V_ELP_R indicates the correlation for lexical decision (i.e., LexicalD) time (i.e., RT) of visual modality (i.e., V) from ELP. Similar naming convention for other variables

Correlations between each independent variable and dependent variables for nonwords

For nonwords, the correlations between each of four reaction time measures with a set of independent variables showed

that NLett was the most significant predictor among all of the predictors irrespective of visual or auditory lexical decision times (Fig. 11). Orthographic uniqueness point measures were also informative predictors, with the exception of the MALD database. The contribution of unigram, bigram, trigram frequencies or OLD20 to reaction time measures varied across databases. The overall correlations for MALD were the weakest among all of the databases.

The different nonword datasets are largely non-overlapping. To compare datasets, analyses were conducted on
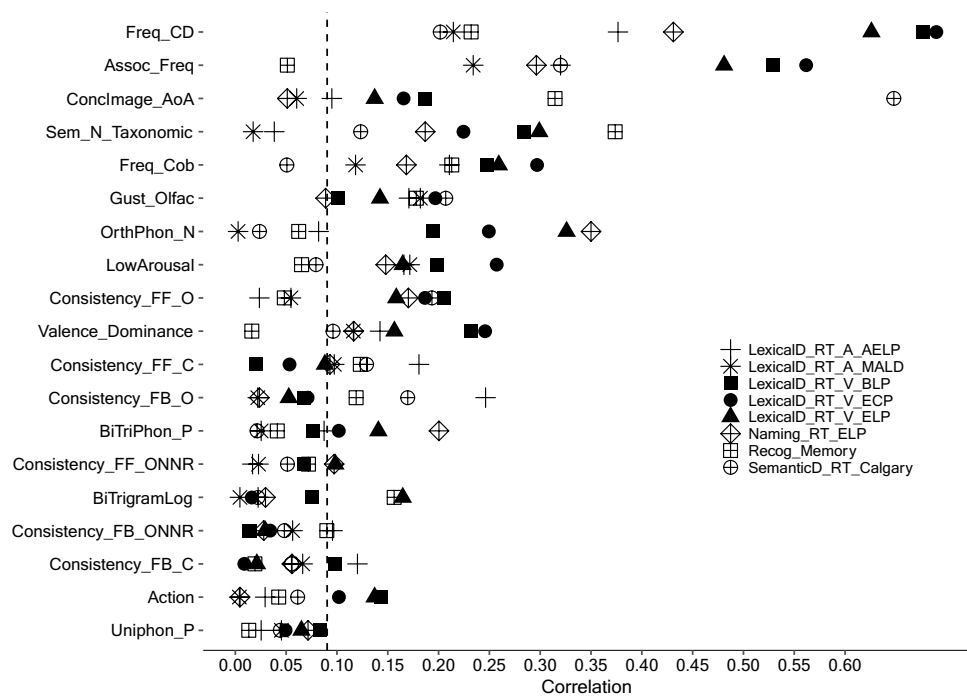
distributional vectors had larger contributions to the dependent variables overall than the best individual variables such as contextual diversity and frequency.

Fig. 10 Absolute correlation values for 19 factors with each of the eight dependent variables (seven zRTs and d' for recognition memory) over 471 words. The variables are ordered in a descending order by overall weighted absolute correlation values. The *vertical dashed line* indicates a threshold for $p < 0.05$ level of significance

Discussion and conclusion

We presented a curated integration of psycholinguistic databases in the form of a metabase, to create the most comprehensive psycholinguistic database to date. The metabase is accompanied by a web interface (https://go.sc.edu/scope/), in which users can either obtain variable values for a given list of words/nonword or generate words/nonwords based on variable values within a range. Our primary goal here was to present the database, rather than answer any specific psycholinguistic questions. Nonetheless, we present some observations from the preliminary analyses below. We conducted two kinds of analyses, one examining the organization or clustering within the variables, and the second related to the correlation between dependent and independent variables.

overlapping portions of pairs of datasets (Fig. 12). The correlations were most consistent between ELP and BLP, while those between visual/auditory datasets were less consistent as expected. The contribution of length-related variables i.e., NLett and OUP, were the most significant predictor to lexical decision times regardless of different samples of overlapping words. Unigram and trigram frequencies were the next most predictive variables, but in opposite directions.

The analyses on the interrelations between a large set of variables showed that variable groupings were generally consistent with theoretical categories (General, Orthographic, Phonological, Semantic, Orth-Phon, Morphological). Variables within the same categorical assignment (e.g., General) were more likely to group together, as expected. Among the clusters, the analyses consistently showed that semantic variables were relatively more spread out than general, orthographic or phonological variables. This is not surprising given that variables related to semantics are generally more complex and subjective (defined by observer, e.g., valence of a word), compared to general, orthographic or phonological variables (defined by wordforms themselves, e.g., orthographic length of a word). Moreover, network analyses indicate that even different types of semantic variables – sensory/perceptual, affect, polysemy and semantic neighborhood related – are distinct in their characteristics and do not cluster together.

The network analyses showed that morphological frequency and orthographic length were the variables that most strongly connected to other variables. These findings are consistent with previous evidence suggesting importance of morphology in the representation of the lexicon (Caramazza et al., 1988; Kuperman et al., 2008). Orthographic length is well-known to have an effect at an early temporal stage of word processing (e.g., Hauk et al., 2006), which suggests

Table 5 Actual correlation values for 19 factors with each of the eight dependent variables (seven zRTs and d' for recognition memory) over 471 words

Factor_ID	FactorNames	Weighted_Overall_R	LexicalD_RT_V_ELP_R	LexicalD_RT_V_ECP_R	LexicalD_RT_V_BLP_R	LexicalD_RT_A_MALD_R	LexicalD_RT_A_AELP_R	Naming_RT_ELP_R	Recog_Memory_R	SemanticD_RT_Calgary_R
PA1	Freq_CD	-0.365	-0.626	-0.69	-0.676	-0.214	-0.377	-0.431	-0.232	-0.202
PA11	Assoc_Freq	-0.294	-0.481	-0.562	-0.529	-0.234	-0.32	-0.296	0.051	-0.32
PA3	ConcImage_AoA	0.251	0.137	0.166	0.187	0.06	0.095	0.051	-0.315	0.648
PA7	Sem_N_Taxonomic	-0.196	-0.299	-0.225	-0.285	0.017	-0.038	-0.187	-0.374	0.123
PA12	Freq_Cob	-0.173	-0.259	-0.297	-0.248	-0.118	-0.211	-0.168	-0.213	0.051
PA16	Gust_Olfac	-0.159	-0.142	-0.197	-0.101	-0.182	-0.171	-0.089	0.178	-0.207
PA2	OrthPhon_N	-0.147	-0.326	-0.249	-0.195	-0.003	-0.082	-0.35	-0.062	-0.024
PA19	LowArousal	0.134	0.165	0.257	0.199	0.172	0.166	0.148	-0.065	0.079
PA18	Consistency_FF_O	-0.127	-0.158	-0.187	-0.205	-0.055	-0.024	-0.17	-0.048	-0.193
PA8	Valence_Dominance	-0.114	-0.157	-0.246	-0.232	-0.116	-0.142	-0.117	0.016	-0.096
PA13	Consistency_FF_C	-0.108	-0.088	0.053	-0.021	0.097	0.181	-0.093	-0.123	0.129
PA10	Consistency_FB_O	-0.102	-0.053	0.071	0.068	-0.022	0.246	-0.024	-0.119	0.17
PA17	BiTriPhon_P	0.085	0.141	0.102	0.077	0.025	-0.087	0.2	-0.041	0.021
PA6	Consistency_FF_ONNR	-0.064	-0.098	-0.068	-0.067	-0.023	-0.017	-0.097	-0.072	0.051
PA14	BiTrigramLog	-0.061	-0.165	-0.016	-0.076	-0.005	0.022	-0.03	-0.156	0.022
PA4	Consistency_FB_ONNR	-0.054	-0.029	0.034	0.014	-0.056	0.096	-0.028	-0.09	0.048
PA9	Consistency_FB_C	-0.053	0.021	0.009	0.098	-0.066	-0.12	-0.056	0.019	-0.055
PA15	Action	-0.051	-0.137	-0.102	-0.143	0.004	-0.029	0.005	-0.043	-0.061
PA5	Uniphon_P	0.046	0.065	0.05	0.083	0.045	0.025	0.072	0.013	-0.044

Weighted_Overall_R indicates the overall weighted correlation that gives equal weight to each task (visual lexical decision, auditory lexical decision, naming, recognition memory, and semantic decision). LexicalD_RT_V_ELP_R indicates the correlation for lexical decision (i.e., LexicalD) time (i.e., RT) of visual modality (i.e., V) from ELP. Similar naming convention for other variables
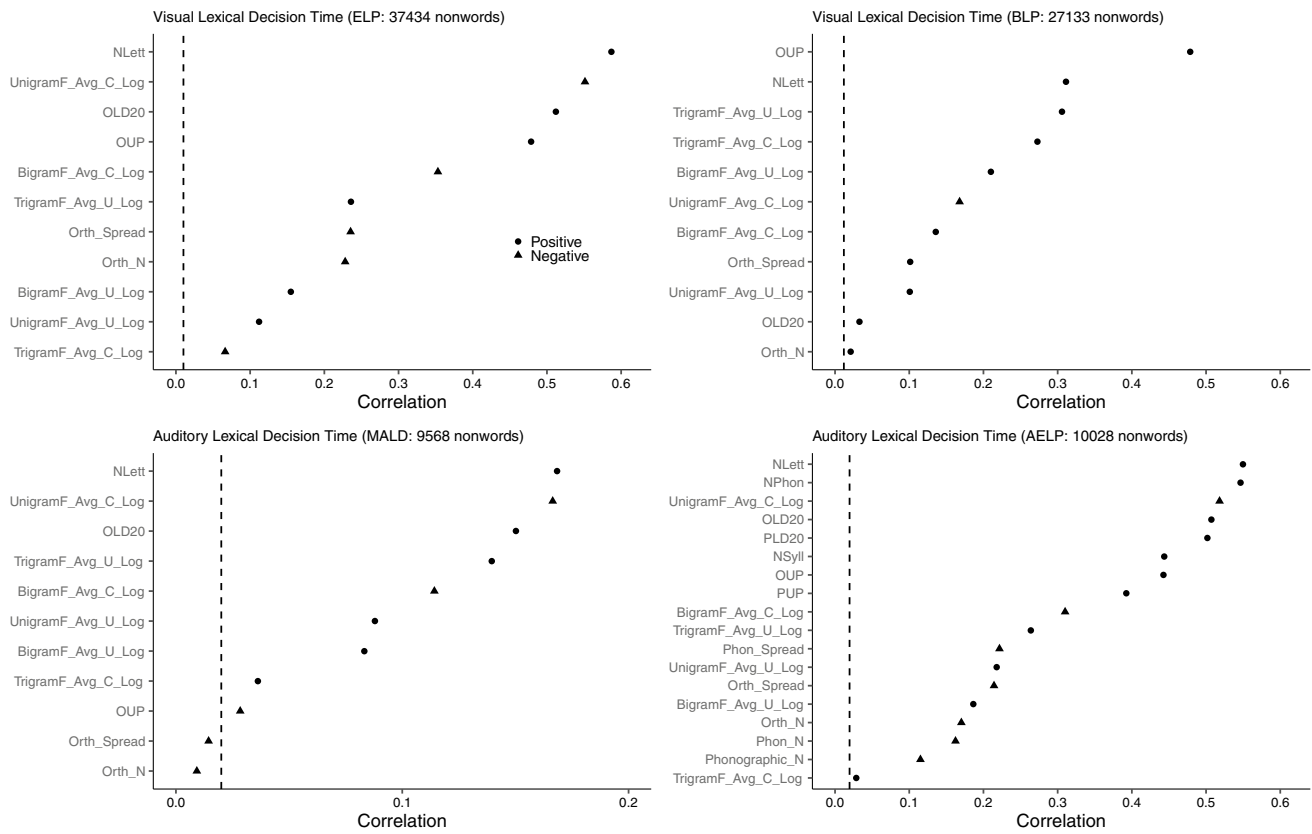
Fig. 11 Actual correlation values of zRT measures for nonwords from ELP, BLP, MALD, and AELP databases with independent variables. The variables are ordered in a descending order by the absolute cor-relation values. The *vertical dashed line* indicates a threshold for $p < 0.05$ level of significance
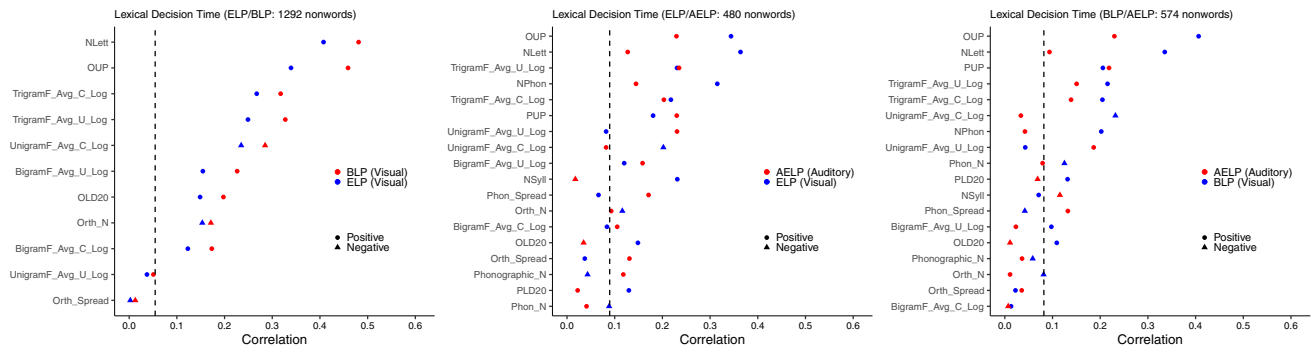

Fig. 12 Absolute correlation values of zRT measures for nonwords with a series of independent variables for overlapping samples between ELP (visual lexical decision time) and BLP (visual lexical decision time) with 1,292 words, ELP and AELP (auditory lexical decision time) with 480 words, and BLP and AELP with 574 words, respectively. The variables are ordered in a descending order by the sum of absolute correlation values. The *vertical dash line* indicates a threshold for $p < 0.05$ level of significance

it may influence other cognitive processes. In addition, we found phonographic neighborhood was the variable most close to other nodes in the network. As a combination of orthographic and phonological neighborhoods (Peereman & Content, 1997), it is suggested to be more important in lexical representations compared to orthographic neighborhood

(Adelman & Brown, 2007). Our results demonstrated a central role of phonographic neighborhood that connects orthographic and phonological neighborhood variables. At the other end, affect was found to have low strength, low betweenness, and low closeness. Thus, affective attributes of words appear to be captured by other variables in the

network, at least for the Warriner et al. (2013) measure that was selected.

The analyses on the correlations between variables/clusters and the dependent variables showed that overall contextual diversity and frequency variables had the largest correlation with the dependent variables. This replicates many previous results (Adelman et al., 2006; Brysbaert et al., 2018; Monsell et al., 1989), but on an unprecedented scale in terms of the number of dependent and independent variables examined. Overall, CD/frequency, association frequency, AoA, and taxonomic semantic neighborhood were the strongest factors across tasks.

The changes in the ranking and correlation of variables due to the change in dependent variable are instructive. The CD/frequency factor is far and away the strongest predictor for the visual lexical decision task. This indicates the importance of exposure and familiarity of the surface from for visual lexical decision. For auditory lexical decision, the results are very different, in that overall correlations are much lower. Not many variables other than CD/frequency, AoA, and association frequency have strong correlations for auditory lexical decision. There are also significant differences between AELP and MALD, with MALD correlations being especially weak. For naming, no one variable or factor dominates. CD/frequency, association frequency, orthographic neighborhood and Orth-Phon variables have very similar strengths. This is consistent with previous evidence that demonstrated the importance of phonographic variables in naming (Adelman & Brown, 2007).

For recognition memory and semantic decision, semantic factors come to the fore. Concreteness/imageability and taxonomic semantic neighborhood were the strongest for recognition memory. Coding of items in memory is strongly reliant not just on being able to form an image of the item, but also on the number of (taxonomically) similar items. For semantic decision, concreteness/imageability is the strongest factor by far, and nothing else comes close. The second most important factor of association frequency has less than half the correlation compared to concreteness/imageability, which is noteworthy even given the fact that semantic decision task explicitly required judging concreteness (Pexman et al., 2017). In contrast to the memory task, the semantic neighborhood factor has a somewhat lower correlation for semantic decision. For a semantic task, the sensory features of the item itself are primarily relevant, and the effects of spreading activation in a semantic neighborhood come into play only in the context of a memory task. As opposed to the lexical decision and naming tasks, CD/frequency has a significant but much lower importance for both memory and semantic decision tasks, setting up a contrast between the value of exposure to the surface form vs. access to sensory features. Perhaps surprisingly, the gustatory/olfactory semantic factor had strong correlation with recognition

memory and semantic decision, but with no other tasks. On the other hand, both association frequency and AoA strongly predicted all tasks except recognition memory. These results underscore the fact that these tasks, including visual and auditory lexical decision, rely on significantly different psycholinguistic processes, and are not interchangeable. No one task can be taken as a standard index of "word processing."

Among consistency variables, the feedforward onset consistency was the most correlated with dependent measures, with a strong correlation with not only for visual lexical decision and naming tasks, but also for semantic decision. This can be related to the debate between single- vs. dual-pathway models of reading (Seidenberg, 2012). The single-system view has argued that a semantic pathway is used to read inconsistent words, and the ability for consistency to predict semantic (concreteness) decision times appears to support this view.

We especially draw attention to the taxonomic semantic neighborhood factor, introduced by Reilly and Desai (2017), which is novel to SCOPE and has been rarely used in psycholinguistic research. It had a strong correlation with all dependent measures, with the exception of auditory lexical decision tasks. It was the second strongest variable predicting recognition memory. Association frequency is another factor that had strong correlations with all tasks, but is not commonly used. These results suggest that these two variables can become part of a standard set of psycholinguistic covariates, along with popular variables such as frequency, length, concreteness/imageability, and age of acquisition.

We found that the distributed semantic vectors consistently outperformed all other individual variables in predicting the dependent variables across visual and auditory lexical decision, naming, recognition memory, and semantic decision, which is a novel result to our knowledge. Previous studies have shown that such distributed semantic vectors and can be used to predict human performance in a range of tasks such as word associations and similarity judgments (Landauer & Dumais, 1997; Pereira et al., 2016). Our findings highlight the promising aspects of distributional semantic vectors in representing word meanings. We found that Word2Vec, GloVe, and Taxonomic distributional vectors have comparable performance in predicting a range of tasks. A current debate pertains to the difference between distributional semantic vectors derived purely from statistical co-occurrence patterns in text corpora, and those derived from experiential attributes, with respect to capturing underlying semantics of words. Some recent neuroimaging results suggest an advantage for experience-based vector representations (Fernandino et al., 2022). Here, we were not able to directly compare distributional vectors to experiential attributes (Compo_attribs in this database) due to the relatively small size of the latter. A future direction is to increase the size of the experiential attributes set and compare their

ability to predict these behavioral dependent measures with those of the three distributional vectors.

For nonwords, length and uniqueness points were found to have the highest correlation with dependent measures. Unigram, trigram, and bigram frequencies followed in their predictive value, with trigram frequencies ranking high, in contrast to the results for words, where trigram frequencies ranked lower than bigram and unigram frequencies. High unigram frequency was faciliatory, while high trigram frequency increased latency. This is consistent with the intuition that word-likeness of nonwords, indexed by trigram frequency, is an important factor for determining their latencies. Neighborhood measures such as OLD20 had a weaker but significant effect on nonword processing times. Orthographic spread was surprisingly found to have no significant correlation with dependent measures, suggesting that this factor does not play a major role in word processing, even without factoring out covariates such as frequency and length.

The results (Supplemental Tables 3 and 4) can be used to pick the "best" measures, and select among alternative measures of a nominally same variable, given the overall weighted correlation as well as correlation for each dependent variable. For example, we found that frequency measures from Worldlex (Gimenes & New, 2016) and SUBTLEX (Brysbaert & New, 2009; Van Heuven et al., 2014) datasets had generally the strongest correlations with dependent variables compared to other frequency measures. We note that the CELEX (COBUILD) frequencies cluster separately from all other frequency measures when using any clustering method (t-SNE, hierarchical clustering, or factor analysis) and have lower correlations with dependent measures. This indicates a qualitative difference in corpus characteristics, and suggests that other frequency measures may be more suited for psycholinguistic research.

No large differences in the overall weighted correlations across a range of tasks were found between measures for many variables such as contextual diversity, age of acquisitions, concreteness etc. The identification of factors underlying this large group of variables (Table 4) may be used to guide future research. For example, picking a representative variable for each factor may be more desirable than picking an arbitrary or customary set of "standard" variables.

The rankings of the variables obtained in the correlation analyses should be interpreted with caution for several reasons. First, the precise order can change depending on the specific words selected or the dependent measure. Nonetheless, we expect that the pattern of relative importance of various factors for various tasks should remain stable even with different word sets. Second, the contribution of the variables to dependent measures also depends on how much variance they explain over and above other variables. Because we only examined each variable in a univariate

manner, interactions between variables were not explored. For example, the frequency × consistency interaction in word naming is well known (Seidenberg et al., 1984), where consistency has a low effect for high frequency words, but a strong effect for low frequency words. Such effects are not seen in the current analyses, resulting in consistency being rated relatively low, which is arguably misleading with regard to the importance of this variable. Such theoretically relevant interactions can be explored in future studies. Future studies can also investigate which cohort of variables explains the most variance as a group, in both linear and nonlinear models. Third, differences in tasks instructions may affect what an individual variable captures. For example, for association frequency measures, some studies elicited free associations in the broad possible sense; whereas some other studies asked participants to give a *meaningful* response which may affect what responses were given (see De Deyne et al., 2019 for a discussion; Nelson et al., 2004). Similarly, for concreteness ratings, some studies emphasize visual properties, while others do not.

While we have attempted to be expansive in coverage of variables that are used in psycholinguistic research, we have also not replicated all megastudy databases in their entirety. Many of the databases that we have integrated contain some unique features or variables that are not included here. For example, the AELP database (Goh et al., 2020) contains multiple auditory recordings of words and nonwords that are not included here. The Lancaster Sensorimotor Norms (Lynott et al., 2020) provides a number of summary variables, such as Minkowski 10, Minknowski 3, Summed Strength, and Max Strength. We have only included the Minkowski 3 measure, as it was found to be the best measure for predicting lexical decision response times and accuracy. It is conceivable that other summary measures might be useful in other circumstances. We direct users to the original database for the full set of variables and features, and hope that SCOPE will serve as a portal for discovery of new and informative variables. We have not included commercial databases such as CoCA (Corpus of Contemporary American English), which would not allow free sharing of the data. The metabase is currently restricted to words and nonwords. For example, the current version does not include picture stimuli that are commonly used in object, verb, famous face, or landmark naming tasks. SCOPE also does not contain multi-word combination or sentence-level norms, or norms that pertain to two or more specific words (e.g., association strength between two words). Included variables are norms that are calculated from the wordform alone, or from the wordform and a dictionary. An important limitation of the current version is also that the metabase is restricted to English. Future versions may expand it to other languages.

The data can be freely explored or downloaded from a web interface and search engine (https://go.sc.edu/scope/).

We have attempted to make the interface user-friendly, to make it easy to select variables, obtain variable values from a given list of words/nonwords, and generate words/nonwords based on variable values within a range. The back end of the metabase is also designed such that addition of new variables is not cumbersome, as development of new variables is inevitable. We hope that the ease of use and continued updates will promote the development of improved psycholinguistic models and facilitate a better understanding of the contribution of these variables, their interactions, and tasks to processing of language. We also hope that the metabase will help standardize practice in psycholinguistics and related disciplines.

Data availability The data for the present study can be accessed at go.sc.edu/scope.

Code availability The code for the present study can be accessed at https://osf.io/9qbjz/.

Declarations

Conflicts of interest None.

Ethics approval This study does not involve any data collection; therefore, no ethics approval is needed.

Consent to participate This study does not involve any data collection; therefore, no consent to participate is needed.

Consent for publication All authors approve for this publication.

References

Adelman, J. S., & Brown, G. D. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*(3), 455–459.

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*(3–4), 130–174.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33*(1), 73–79.

Brysbaert, M. (2017). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology, 70*(7), 1129–1139.

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.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*(4), 991–997.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods, 51*(2), 467–479.

Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods, 51*(4), 1849–1863.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*(3), 297–332.

Chee, Q. W., Chow, K. J., Goh, W. D., & Yap, M. J. (2021). LexiCAL: A calculator for lexical variables. *Plos One, 16*(4), e0250891.

Chee, Q. W., Chow, K. J., Yap, M. J., & Goh, W. D. (2020). Consistency norms for 37,677 English words. *Behavior Research Methods, 52*(6), 2535–2555.

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371–383.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In *Attention and performance VI* (pp. 535-555). Routledge.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384–387.

Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70*(6), 885–901.

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary, the words we know: A national vocabulary inventory*. World book .

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*(3), 987–1006.

Diveica, V., Pexman, P. M., & Binney, R. J. (2022). Quantifying social semantics: An inclusive definition of socialness and ratings for 8388 English words. *Behavior Research Methods.* https://doi.org/10.3758/s13428-022-01810-x

Engelthaler, T., & Hills, T. T. (2018). Humor norms for 4,997 English words. *Behavior Research Methods, 50*(3), 1116–1124.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212.

Fernandino, L., Tong, J. Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences of the United States of America, 119*(6). https://doi.org/10.1073/pnas.2108091119

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427.

Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior Research Methods, 48*(3), 963–972.

Goh, W. D., Yap, M. J., & Chee, Q. W. (2020). The Auditory English Lexicon Project: A multi-talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods, 52*(5), 2202–2231.

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition. *Frontiers in Psychology, 5*, 1307.

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral cortex, 20*(8), 1799-1815.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage, 30*(4), 1383–1400.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.

Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods, 45*(1), 160–168.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287–304.

Khanna, M. M., & Cortese, M. J. (2021). How well imageability, concreteness, perceptual strength, and action strength predict recognition memory, lexical decision, and reading aloud performance. *Memory, 29*(5), 622–636.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language & Cognitive Processes, 23*(7–8), 1089–1132.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211.

Levenshtein, V. I. (1966). *Binary codes capable of correcting deletions, insertions, and reversals*. Soviet Physics Doklady.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203–208.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: Multidimensional measures of Perceptual and Action Strength for 40,000 English words. *Behavior Research Methods, 52*(3), 1271–1291.

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods, 52*(2), 741–760.

Medler, D.A., & Binder, J.R. (2005). *MCWord: An on-line orthographic database of the English language*. http://www.neuro.mcw.edu/mcword/

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Mohammad, S., & Turney, P. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text.

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence, 29*(3), 436–465.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118*(1), 43.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification, 31*(3), 274–295.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1p2), 1.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language, 37*(3), 382–410.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology, 33*(3–4), 175–190.

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words. *Behavior Research Methods, 49*(2), 407–417.

Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods, 51*(2), 453–466.

Reilly, M., & Desai, R. H. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition, 169*, 46–53.

Rice, C. A., Beekhuizen, B., Dubrovsky, V., Stevenson, S., & Armstrong, B. C. (2019). A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings. *Behavior Research Methods, 51*(3), 1399–1425.

Roller, S., & Erk, K. (2016). Relations such as hypernymy: Identifying and exploiting Hearst patterns in distributional vectors for lexical entailment. *arXiv preprint arXiv:1605.05433*.

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods, 50*(4), 1568–1580.

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*(3), 1258–1270.

Seidenberg, M. S. (2012). Computational models of reading: connectionist and dual-route approaches. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *Cambridge Handbook of Psycholinguistics* (pp. 186–203). Cambridge University Press.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*(3), 383–404.

Shaoul, C., & Westbury, C. (2006). Word frequency effects in high-dimensional co-occurrence models: A new approach. *Behavior Research Methods, 38*(2), 190–195.

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413.

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods, 52*(6), 2372–2382.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Lawrence Erlbaum.

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior Research Methods, 51*(3), 1187–1204.

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine phonotactic online dictionary*. Version 2.0. [Data file].

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*(1), 183–190.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374–408.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207.

Weide, R. (2005). *The Carnegie Mellon pronouncing dictionary [cmudict. 0.6]*. Carnegie Mellon University.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971–979.