OPEN ACCESS

Vocabulary Learning and Instruction

Volume 12, Number 1 (2023)
https://doi.org/10.29140/vli.v12n1.1055

Castledown

Semantic Variables and their Application in L2 Research

Allie Spencer Patterson
Rikkyo University, Japan
tuf57145@temple.edu

Abstract

Semantic variables enable L2 researchers and materials creators to quantify and control the effects of meaning on cognition. However, in recent years, many variables have been normed and published. Parsing the methods employed in norming this myriad of variables and which disparate theories informed their creation can be an opaque and arduous task. To facilitate effective use of these measures, this study consists of a literature review of concreteness, imageability, semantic network variables, and embodied variables. In each section, the theory underlying each variable is first outlined with the methods employed in norming studies delineated next. The final part of each section consists of an exemplary study in which the semantic variable was employed in analyses. The overarching goal of this review is to facilitate effective theory-driven use of these variables in L2 research and materials creation.

Keywords: semantic, imageablility, concreteness, embodiment, semantic network, computationalism

Introduction

Semantics play a key role in lexical processing and learnability (Schmitt & Schmitt, 2020). In pursuit of quantifying the effect of meaning on language processing, linguists have developed a variety of semantic variables which have often been applied in both

Data Availability Statement: All relevant data are within this paper.

L2 research and materials development. However, how researchers use a semantic variable depends on the background theory informing its development and the norming procedures employed in its creation. Due to the growing number of available semantic variables, L2 researchers may have difficulty parsing these factors. Choosing an appropriate semantic variable for a research project or for L2 materials creation can be a daunting task due to the considerable number of available options.

In consideration of the above issues, the current study is a literature review meant to facilitate the effective use of semantic variables in L2 research and pedagogy. The overarching goal of this review is to provide a guide for L2 researchers who are attempting to identify which semantic variable is best for their study or material creation. In pursuit of this goal, this study consists of a review of current literature on concreteness, imageability, semantic network variables, and embodied variables. First, the underlying theory behind each variable is discussed, followed by an outline of the norming methods used to develop the measures. Finally, an exemplary study is reviewed in detail to contextualize the appropriate use of these variables in research.

Concreteness

Concreteness, the oldest of the semantic variables, was developed from the early 21st century findings of Jung (1919) and Woodworth and Schlosberg (1938) who demonstrated that participants could process and respond to words associated with real-world sensory experiences faster than those referring to abstract concepts. For instance, participants would process *apple* or *pencil* faster than *honor* or *death*. Faster responses to a prime in comparison to another indicate that the faster item required less processing (Jiang, 2013). Early formalized lists of concreteness ratings were developed in the mid-20th century (Darley et al., 1959; Gorman, 1961), but the modern iteration of the variable arose in the late 1960s (Paivio et al., 1968; Spreen & Schulz, 1966).

In concreteness item norming studies, researchers recruit participants to rate single words on a seven-point Likert scale based on an item's "sense-experience" (Spreen & Schulz, 1966, p. 460). The participants are instructed to rate abstract words not directly tied to sensory experiences low and to relate concrete words associated with sensory experiences high. Spreen & Schulz (1966) gave participants the below example to contextualize their rating:

> Think of the words "chair" and "independence." "Chair" can be experienced by our senses and therefore should be rated as high concrete; "independence" cannot be experienced by the senses as such and therefore should be rated as low concrete (or abstract). (p. 460).

The researchers averaged the participant's rating obtained with these instructions and presented the results as tables within their study. An example of a highly rated concrete item from Paivio et al. (1968) is *forehead* with a mean of 6.93, whereas an example of a low rated item abstract item is *idea* with a mean of 1.42. Analogous Likert scale norming methods are employed for many of the variables discussed in this review.

In recent years, the rating instructions have evolved from a focus on sense-experience to a focus on reference to an external entity. For instance, Brysbaert et al. (2014) had participants rate "the degree to which the concept denoted by a word refers to a perceptible entity" (p. 904). In addition, the research had participants rate words on a Likert scale of 1 to 5 instead of the seven-point scale. For example, *forehead* in Brysbaert et al. has a mean rating of 4.5, and *idea* has a rating of 1.61. These methodological changes may have been due to concreteness being functionally indistinguishable from imageability (Reilly & Kean, 2007) and the variable relying too heavily on visual perception (Connell & Lynott, 2012) to the detriment of other sense modalities.

Despite existing for more than sixty years, researchers continue to publish concreteness ratings and apply them in the study of linguistic processing. Due to the ease of collecting data through crowdsourcing, researchers are now able to obtain massive datasets of ratings, such as Brysbaert's et al.'s (2014) ratings for 40,000 English words and Muraki et al.'s (2022) database of rating for 62,000 multiword expressions. Ratings for languages other than English are available, such as Spanish ratings in the EsPal database (Duchon et al., 2013) or Croatian ratings (Ćoso et al., 2019). Recent novel applications of concreteness include using the ratings as a means of comparing story telling efficacy between two neural network large language models (See et al., 2019) and as a dependent variable which predicts the ability of a noninvasive brain sensor to reconstruct processed language from fMRI (functional magnetic resonance imaging) readings (Tang et al., 2023). Concreteness has been demonstrated to be a factor in ease of L2 multi-word item retrievability (Lindstromberg & Eyckmans, 2022), and to inform item selection in a study of semantic fluency (Fernández-Fontecha, 2021). Despite recent criticism of the construct (Barsalou et al., 2018), linguistics researchers will likely continue to utilize concreteness ratings as a means of exploring semantic effects.

Concreteness Exemplar: Vitta et al. (2023)

To explore which lexical characteristics predict the probability an L2 learner will recognize a word, Vitta et al. (2023) administered a yes-no test to Arabic and Japanese L2 learners of English. In this testing format, participants indicated whether they do or do not know a word by clicking yes or no. Intermixed within the target real word items are pseudowords to control for participant performance. This study was a conceptual replication of Hashimoto & Egbert's (2019) yes-no test study. However, Vitta et al. used more conservative methods than the original study. Firstly, in contrast to Hashimoto & Egbert, the researchers employed a theory driven approach to the construction of regression models instead of stepwise regression. Stepwise regression model construction has been criticized for not being sufficiently hypothesis and theory driven, and instead allows for model creation through mathematics alone with little theoretical input from the researcher (Smith, 2018).

Secondly, Vitta et al. controlled participant performance through the inclusion of false alarm rate covariate in their second model. This variable is the rate at which a participant answered yes to indicate they knew pseudowords. Inclusion of the false alarm rate in models enabled the researchers to control for non-target variance attributable to a participant's propensity to falsely indicate they knew a word. The probability of false positives could influence task performance not only for the pseudowords but also on the target real words. As such, the inclusion of this covariate provides a more holistic and empirical account of which lexical variables affect word recognition. In Hashimoto & Egbert, positive pseudoword responses were only used to remove data from problematic participants who indicated they knew the pseudowords at a rate higher than 10%. In keeping with their theory driven approach, Vitta et al. identified concreteness as a variable of interest using the *Tool for the Automatic Analysis of Lexical Sophistication* (TAALES 2.0) application (Kyle et al., 2018) which analyzes words and identifies variables which may affect lexical complexity. Vitta et al. conducted two iterations of their model, both of which contained concreteness scores as a dependent variable. In the first model which did not contain the false alarm covariate rate, concreteness did not significantly predict word recognition probability. However, when the researchers controlled for participant task performance with the covariate, concreteness was significant.

Vitta et al. was chosen as the exemplar for concreteness for two reasons. Firstly, choosing a semantic variable for an experiment or for materials creation should be theory driven. In contrast to Hashimoto & Egbert's (2019) stepwise approach, the researchers identified concreteness as a variable of interest using prior theory reviewed in Kyle et al. (2018). Secondly, Vitta et al.'s study demonstrates the difficulty of quantifying semantic effects. Concreteness significantly predicted test performance only after the false alarm covariate was included in the model. Unlike the robust effects of frequency, semantic effects may only be observable when mediating confounds have been controlled for. As such, the study of semantic effects requires diligence on the part of the researchers. This requirement is especially true in the study of L2 semantic effects, as the study of SLA presents additional item, participant, and contextual level confounds not present in L1 research.

Imageability

Imageability is an additional rating derived semantic variable developed almost fifty years ago by Paivio (1965) who conceived this variable as a means of supporting research into dual coding theory which proposed that semantic representations are grounded in verbal associations and specific sensorimotor systems (Paivio, 1991). In contrast to modern embodiment theories which will be discussed below, in dual coding theory, only visual, auditory, and motor systems constitute a portion of a lexical item's semantic representation in conjunction with word associations. Dual coding theory continues to be the foundation of some modern SLA research (e.g., Kanellopoulou et al., 2019; Morett, 2019; Wong & Samudra, 2021).

In imageability norming studies, participants are asked to rate "the ease or difficulty with which [words] arouse mental images" (Paivio et al., 1968, p. 4). Participants rate individual words on a Likert scale of one to seven with one indicating difficulty creating a mental image and seven indicating ease of creating a mental image. Scores from all participants are averaged and presented along with the standard deviations. Unlike concreteness, methods for acquiring imageability scores have not undergone any significant revision since early studies beyond adapting rating instructions from paper booklets to computer-based questionnaires (Cortese & Fuggett, 2004; Gilhooly & Logie, 1980; Su et al., 2022).

Linguistics researcher widely use imageability, and the facilitatory effects that high imageability have on comprehension has been documented using a wide range of methods from reaction time studies (Yap et al., 2012), to brain imaging (Wise et al., 2000) and brain lesion studies (Dubé et al., 2014). Recent novel uses of the variable include using neural networks to predict the imageability ratings of words (Bochkarev et al., 2021) and as a means of gleaning insights into the speech patterns of children with autism spectrum disorder (Lin et al., 2022). SLA researchers also often use imageability as an independent variable to predict L2 performance (e.g., Farley et al., 2012; Hasegawa, 2010). For instance, Ellis and Beaton (1993) found that highly imageable vocabulary was more easily learned by L2 students. Research has shown that the use of less imageable words in speech increases as L2 proficiency increases (Salsbury et al., 2011). Researchers have also found that imageability is a statistically significant predictor of ratings of comprehensibility of L2 speech (Saito et al., 2016).

Imageability & Concreteness

Though concreteness focuses on the degree to which a word references a perceptible entity and imageability is associated with mental imagery, the two variables are strongly correlated. For instance, in a comparison of semantic variables, Khanna and Cortese (2021) found a high correlation between concreteness and imageability ($r = .79$) for their chosen items. Since the development of these two variables, their intricate relationship has been often discussed and debated (e.g., Clark & Paivio, 2004; Friendly et al., 1982; Paivio et al., 1968) with most academics arguing for the variables' synonymy (e.g., Connell & Lynott, 2012; Khanna & Cortese, 2021; Reilly & Kean, 2007). Researchers who adopt this stance argue that imageability and concreteness are interchangeable and which variable a researcher chooses for their study often seems to depend on convenience and not theory.

However, some academics have argued that functional distinctions exist between concreteness and imageability. The distribution of concreteness scores is bimodal in contrast to the unimodal distribution of imageability (Kousta et al., 2011; Nelson & Schrieber, 1992). This bimodal distribution indicates that raters do not perceive concreteness as a continuum, but instead view it as a concrete/abstract dichotomy. This contrast arises from instructions in norming studies, as concreteness raters are tasked with indicating the degree that a word indicates a perceptible entity. As such, concepts with strong relationships to sensorimotor experiences, but which do not indicate a physical object, receive low concreteness ratings. In contrast, imageability's norming studies require participants to focus on ease of creating mental images which allows for higher ratings of sensorimotor related concepts with no relationship to physical objects. For instance, the concreteness rating of *freedom* is 277, but its imageability rating is 437 in the MRC Psycholinguistics database (Coltheart, 1981). Due to *freedom* indicating no singular object, the concept receives a low concreteness score. However, *freedom*'s relationship to introspective and affective experiences despite being abstract facilitates the creation of mental imagery demonstrated in its higher imageability rating.

Despite the distinctions discussed above, researchers should avoid using both imageability and concreteness in regression analyses due to multicollinearity

(Daoud, 2017). Inclusion of these highly correlated variables in a regression analysis could lead to increased standard errors raising the probability of type II errors (i.e., erroneous non-significance). Again, choosing which to employ in a study should be driven by theory. If an abstract/concrete dichotomy is necessary, concreteness is most appropriate. Researchers should choose imageability if a unimodal distribution of ratings based on a variety of sensorimotor experiences is more appropriate. If deemed appropriate, both variables can be used to inform L2 pedagogical material creation, if the two variables are not coanalyzed in statistical comparisons.

Imageability Exemplar: Anible (2020)

Anible (2020) examined the relationship between imageability and L2 proficiency effects with L1 English participants whose L2 was American Sign Language (ASL). The experiment was a reaction time translation recognition test. In this type of experiment, participants see a word followed by another, one in their L1 and another in their L2, and must decide if the two are equivalent. In this study, participants were first played a video of an ASL word followed by an audio recording of an English word. Reaction time was measured from the beginning of the video until the participant responded after hearing the recording. 48 correct and 48 incorrect items were included in the experiment, with the incorrect items being the critical target items under analysis. The incorrect items were split into four conditions: unrelated signs, phonologically related signs, semantically related signs, and diagrammatic related signs (i.e., semantically and phonologically related signs). Imageability significantly predicted reaction time across all conditions. Specifically, high imageability was related to interference effects in the phonologically related conditions, and low imageability led to interference effects in the semantic condition. In the diagrammatic condition, low proficiency participants displayed interference effects with low imageability words, but the inverse was found for high proficiency participants, with high imageability being associated with interference effects.

Anible (2020) exemplifies the most typical use of semantic variables which is controlling for semantic effects by creating matched lists of words between conditions. Unlike Vitta et al. (2023) in which concreteness freely varied, Anible controlled for word frequency, word length, concreteness, and imageability across the four conditions. The imageability ratings across each condition sit around an average rating of 500 with a typical standard deviation of around 75 points. This controlled approach should be employed if semantic variables are not of theoretical importance to a study's hypotheses as semantic effects can introduce confounds. Even when researchers control for semantic effects, introducing a semantic variable into models as a covariate can produce fruitful results as is the case with Anible (2020). Despite employing tightly controlled groups across conditions, imageability predicted participants' reaction time across conditions and provided insight into how proficiency mediates reliance on semantics.

Semantic Network Variables

Semantic network variables are derived from the psycholinguistic conceit that when a concept is activated in the mind whether for creating output or processing input, other semantically related concepts are also activated and that a concept's semantics arise in part from these connections (Hameau et al., 2019; Roelofs, 2018). Unlike the other variables discussed in this study, semantic network variables are not grounded in a sensorimotor centric theory of cognitions, such as dual-coding theory or embodied cognition. Instead, researchers developed semantic network variables from computationalist theories (Piccinini, 2009) which is a broad category of sometimes competing theories which attempt to explain the mechanics of cognitions. The theoretical concept most relevant to this group of variables is the connectionist concept of semantic networks (Shirai, 2018). Figure 1 below illustrates a simple semantic network for the concept *bird*. The semantics of *bird* in this example arise in part from connections to other concepts, such as *animal* or *ostrich*. When *bird* is activated for processing, the closely connected terms, such as *animal* or *ostrich*, have a high probability of coactivating while more distant words, such as *cat*, have a lower probability of activating.

Unlike imageability and concreteness, semantic network variables are not a single unified variable with analogous norming methods between studies. Instead, researchers have employed a variety of measures. One method is subjective ratings based on the estimated number of concepts in a word's semantic neighborhood (e.g., Blanken et al., 2022; Bormann, 2011; Bormann et al., 2011). In addition, McRae et al. (2005) distinguished between near and distantly related concepts using features by having participants list features of concepts, such as *dog* (has legs, is a pet, has fur), and then summing the frequency of participant reported features. Researchers have also developed an additional subjective participant variable from free association tasks in which norming participants produce the first word that comes to mind when presented with a word (Mirman & Magnuson, 2008; Nelson et al., 2004). For instance, a participant presented with *bird* might produce *feather*, while another produces *fly*. Finally, other measures (e.g., Hoffman et al., 2013; Shaoul & Westbury, 2010) have been developed from corpus using latent semantic analysis which is a measure of contextual variance of words in texts (Landauer et al., 2008). For instance, using Hoffman et al.'s (2013) SemD (semantic diversity) variable, *offline* has low contextual

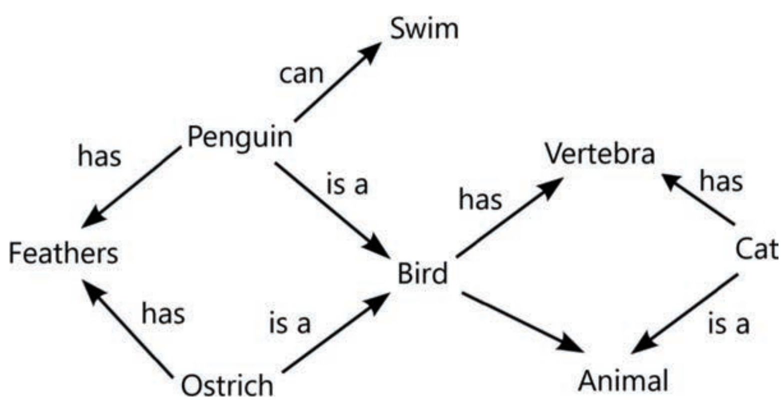

Figure 1 *Bird Semantic Network.*

diversity (.33) meaning it does not appear in many varying contexts in comparison to the contextually diverse concept *addition* (2.24). In consideration of the wide variety of variables, L2 researchers should be mindful of methodologies employed in creating semantic network variables to choose one that best fits the theoretical focus of their studies.

Semantic Network Variables Exemplar: Nenadić et al. (2022)

Though not an L2 study, Nenadić et al. (2022) exemplifies the effective use of semantic network variables by exploring the effect multiple semantic measures have on L1 English listening comprehension. The researchers conducted two auditory lexical decision tasks. In this type of reaction time experiment, participants indicate whether they heard a word and non-words, and their reaction time is measured from the onset of the prime until the participant responds (Jiang, 2013). In the first experiment, researchers replicated an auditory lexical decision task from Goh et al. (2016), but in addition to employing linear regression, the researchers explored the existence of non-linears effects using generalized additive mixed-effects models (GAMMs). In addition to including valence (Kuperman et al., 2014), arousal (Warriner et al., 2013), and concreteness, the researchers explored the effects of two corpus derived semantic networks variables, semantic neighborhood density (Shaoul & Westbury, 2010) and semantic diversity (Hoffman et al., 2013), and one ratings derived variable, number of features (McRae et al., 2005). In the linear regression, valence, concreteness, and number of features were significant, but neither corpus derived semantic network variable significantly predicted participant reaction time. The GAMMs analyses returned analogous results.

In the second study, the researchers expanded the scope of the auditory lexical decision task to include 440 participants completing tasks which contained 8,626 words. This substantial increase in statistical power afforded a more nuanced look into the effect of semantic variables. For instance, in the GAMM analyses, concreteness was found to significantly predict respondent performance for nouns and adjectives, but not for verbs. Concreteness also displayed a non-linear relationship with participants' performance, with ratings from one to four facilitating faster responses as rating increased but items with ratings higher than four having an inhibitory effect on response times. Similarly, semantic neighborhood density also significantly predicted reaction times to nouns and adjectives, but not verbs. Semantic diversity only predicted participant performance for adjectives.

Nenadić et al. (2022) employed three methodological design approaches which should be utilized in future L2 semantic research. First, the researchers employ multiple semantic variables in their analyses. Despite the considerable number of available semantic variables, most studies often only employ one or two in analyses. As such, results do not reflect the interplay between various aspects of semantics. A singular variable cannot quantify the complexity of the human experience of experiencing meaning. Secondly, the use of non-linear analyses, such as GAMMs, presents an opportunity to further insight into the relationships between semantics and cognition. As seen in the example of concreteness, the relationship between semantic variables and cognition is not always linear. Finally, the inclusion of many participants and items boosted the statistical power and generalizability of the second study's results. The proliferation of information technology and data gathering platforms has made data gathering easier than ever before (Patterson & Nicklin, 2023). L2 researchers can utilize these resources with relative ease and without the need for large research budgets to boost the statistical power of their studies.

Embodied Variables

The spread of embodiment theories (Barsalou, 1999; Gallese & Lakoff, 2005; Glenberg, 1997; Pulvermüller, 1999; Zwaan, 1999) in the past twenty years have led to the development of an additional group of semantic variables which quantify the relationship between concepts and sensorimotor experiences. This research paradigm asserts that the foundation of semantics arises from sensorimotor experiences, and semantic processing is in part dependent on processing in sensorimotor brain areas used during perception. At a glance, embodiment seems analogous to dual coding theory (Paivio, 1991) discussed above. However, two principal distinctions exist between the two approaches. Firstly, while language centric embodiment theories have been proposed (e.g., Barsalou et al., 2008; Borghi & Binkofski, 2014; Dove et al., 2022), the core tenets of embodiment are not specific to language, but instead seek to explain the human experience of semantics. Secondly, whereas the dual coding approach asserts that concepts have visual, haptic, and auditory grounding, embodiment researchers go further and argue that all sensorimotor modalities, introspection, and affect ground conceptual semantics (Barsalou, 1999).

Similar to the semantic network variables paradigm, no singular embodiment variable dominates academic discourse. Instead, researchers developed a variety of variables that quantify different sensorimotor, introspective, and affective aspects of concepts. One of the first variables, body-object interaction ratings (BOI; Pexman et al., 2019; Tillotson et al., 2008) quantifies the relationship between a concept's real-world referent and interaction with the human body by having participants rate words on a 7-point Likert scale. For instance, *the* has a mean BOI score of 1.12 due to being abstract and not affording bodily interaction with a real-world referent whereas *toothbrush* has a rating of 6.72. Other variables, such as sensory experience ratings (Juhasz & Yap, 2013), and embodiment ratings (Sidhu et al., 2014) also quantify the relationship between real world referents and bodily interaction. Affective embodiment variables include valence (Kuperman et al., 2014) and arousal (Warriner et al., 2013) ratings. Valence ratings quantify the inherent emotion of a concept, while arousal gauges excitement elicited by a word. In a massive recent study, Lynott et al. (2020) created the Lancaster sensorimotor norms by recruiting 3500 participants to rate approximately 40,000 English words using 5-point Likert scales on five body effectors, six sensory modalities. Table 1 shows the mean ratings for abstract concept *freedom* and the concrete *pen* across the modalities. The researchers also provided holistic variables, the three Minkowski3 variables found at the bottom of Table 1, which incorporate rating across modalities. As of yet, quantitative embodiment research methods have not been widely explored in L2 research. As such, little is known about the predictive ability of this family of variables on L2 performance.

Table 1 *Lancaster Sensorimotor Ratings Example*

	Freedom	Pen
Auditory	2.278	0.526
Gustatory	0.556	0.053
Haptic	1.333	4.316
Interoceptive	3.278	0.053
Olfactory	0.944	0.158
Visual	3.111	4.211
Foot/Leg	3.337	0.238
Hand/Arm	2.136	4.238
Head	2.955	2.286
Mouth	2.545	0.333
Torso	1.955	0.476
Minkowski3 Perceptual	4.318	5.374
Minkowski3 Action	4.132	4.452
Minkowski3 Sensorimotor	5.326	5.326

Embodied Variables Exemplar: Patterson (2022)

One of the few studies to utilize an embodiment semantic variable, Patterson (2022) was concerned with identifying which single word variables affect novice L2 listeners comprehension. 172 novice L2 English participants completed an isolated phrase paused transcription test. In this type of assessment, participants hear a recorded L2 phrase once and must quickly attempt to transcribe it. This study was a continuation of research begun in Patterson (2021) in which frequency, word length, and Minkowski3 sensorimotor norms significantly predicted transcription probability of functor transcription with mid-proficiency English L2 participants. In contrast to this prior study, both functors and content words were included in analyses with part of speech (i.e., whether a word is a functor or content word), phrasal position, word length, frequency, and Minkowski3 sensorimotor norms as independent variables. The participants completed the task using a paper form in their normal L2 classroom, and the results were then coded by the researcher with a 1 indicating successful transcription of the word and a 0 indicating unsuccessful or non-transcription. These dichotomous results were then analyzed using Rasch analysis and hierarchical logit linear mixed effects regression. Part of speech and phrasal position did not significantly predict transcription probabilities. Length, frequency, and Minkowski3 sensorimotor ratings were significant. However, frequency was only significant after the inclusion of the embodiment variable into the model.

Unlike the other exemplar studies and most studies cited in this review which were dependent on computer-based tasks, Patterson (2022) demonstrates that finding semantic effects is possible with intact L2 classes through tasks that approximate typical classroom activities. Developing the paper based listening tests used in Patterson (2021) and Patterson (2022) is possible for any educators with access to a voice recorder and a photocopier. Creating an analogous task is not contingent on access to a computer lab and subscriptions to software. In consideration of this, researchers can and should make use of the semantic variables discussed in this review to aid in material development and in analyses of classroom-based research. Semantics not only affect reaction times but also have an influence on L2 classroom performance and acquisition, such as the ability to recall heard input.

Conclusion

The overarching objective of this review is to make semantic variables more accessible to L2 researchers and educators. In pursuit of this goal, this review delineated the theories and norming methods used to create concreteness, imageability, semantic network variables, and embodiment variables. Each variable discussed in this study employs different norming methods grounded in disparate theories. As such, each variable reflects distinct aspects of semantics. As discussed above, unlike the straightforward measures employed with frequency, a singular all-encompassing variable cannot represent the

Table 2 *Online Semantic Variable Resources*

Concreteness	The tool for the automatic analysis of lexical sophistication (TAALES)
	Brysbaert et al. (2014) – Concreteness ratings for 40 thousand generally known English word lemmas*
	Muraki et al. (2022) – Concreteness ratings for 62 thousand English multiword expressions
	EsPal – Spanish Lexical Variables
Imageability	MRC Psycholinguistics Database
	Su et al. (2022) – Imageability ratings for 10,426 Chinese two-character words*
Semantic Network Variables	Hoffman et al. (2012) – Semantic Diversity - SemD*
	McRae et al. (2005) – Number of Features*
Embodiment Variables	Lancaster Sensorimotor Norms
	Warriner et al. (2013) – Norms of valence, arousal, and dominance for 13,915 English lemmas*
	Pexman et al. (2018) – Body–object interaction ratings for more than 9,000 English words*

Note. * indicates that the database CSV can be found in the online supplementary materials of the study.

intricate experience of human semantics. To make these variables further accessible, Table 2 is a series of links with access to many of the semantic variables and resources discussed in this study. Before employing one of these variables, L2 researchers must consider which variable best fits their pedagogical and theoretical needs.

References

Anible, B. (2020). Iconicity in American sign language–English translation recognition. *Language and Cognition, 12*(1), 138–163. https://doi.org/10.1017/langcog.2019.51

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660. https://doi.org/10.1017/S0140525X99532147

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373,* 1–11. https://doi.org/10.1098/rstb.2017.0144

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0013

Blanken, G., Dittmann, J., Wallesch, C.-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325*, 72–74. https://doi.org/10.1016/s0304-3940(02)00227-6

Bochkarev, V. V., Savinkov, A. V., & Shevlyakova, A. V. (2021). Estimation of Imageability Ratings of English Words Using Neural Networks. In Advances in Soft Computing: 20th Mexican International Conference on Artificial Intelligence, MICAI 2021, Mexico City, Mexico, October 25–30, 2021, Proceedings, Part II 20 (pp. 59–69). Springer International Publishing. https://doi.org/10.1007/978-3-030-89820-5_5

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-9539-0_3

Bormann, T. (2011). The role of lexical-semantic neighborhood in object naming: Implications for models of lexical access. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00127

Bormann T., Kulke F., Wallesch C.-W., Blanken G. (2008). Omissions and semantic errors in aphasic naming: Is there a link? *Brain and Language, 104*(1), 24–32. https://doi.org/10.1016/j.bandl.2007.02.004

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods, 46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, and Computers, 36*, 371–383. https://doi.org/10.3758/bf03195584

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505. https://doi.org/10.1080/14640748108400805

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125*, 452–465. https://doi.org/10.1016/j.cognition.2012.07.010

Ćoso, B., Guasch, M., Ferré, P., & Hinojosa, J. A. (2019). Affective and concreteness norms for 3,022 Croatian words. *Quarterly Journal of Experimental Psychology, 72*(9), 2302–2312. https://doi.org/10.1177/1747021819834226

Cortese, M., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, 36*(3), 384–387. https://doi.org/10.3758/BF03195585

Darley, F. L., Sherman, D., & Siegel, G. M. (1959). Scaling of abstraction level of single words. *Journal of Speech and Hearing Research, 2*(2), 161–167. https://doi.org/10.1044/jshr.0202.161

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series, 949*. https://doi.org/10.1088/1742-6596/949/1/012009

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2022). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research, 86*(8), 2451–2467. https://doi.org/10.1007/s00426-020-01438-6

Dubé, C., Monetta, L., Martínez Cuitiño, M. M., & Wilson, M. A. (2014). Independent effects of imageability and grammatical class in synonym judgement in aphasia. *Psicothema. 26*(4), 449–456. https://doi.org/10.7334/psicothema2014.31

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning, 43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Farley, A. P., Ramonda, K., & Liu, X. (2012). The concreteness effect and the bilingual lexicon: The impact of visual stimuli attachment on meaning recall of abstract L2 words. *Language Teaching Research*, *16*(4), 449–466. https://doi.org/10.1177/1362168812436910

Fernández-Fontecha, A. (2021). The role of learner creativity in L2 semantic fluency. An exploratory study. *System, 103*, 1–14. https://doi.org/10.1016/j.system.2021.102658

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1080 words. *Behavior Research Methods & Instrumentation, 14*, 375–399. https://doi.org/10.3758/bf03203275

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479. https://doi.org/10.1080/02643290442000310

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. https://doi.org/10.3758/BF03201693

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–19. https://doi.org/10.1017/s0140525x97000010

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00976

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*(1). https://doi.org/10.1037/h0040561

Hameau, S., Nickels, L., & Biedermann, B. (2019). Effects of semantic neighbourhood density on spoken word production. *Quarterly Journal of Experimental Psychology, 72*(12), 2752–2775. https://doi.org/10.1177/1747021819859850

Hasegawa, Y. (2010). Context effects on EFL vocabulary learning: Focusing on word imageability and test formats. *JLTA Journal*, *13*, 145–161. https://doi.org/10.20622/jltaj.13.0_145

Hashimoto, B. J., & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning, 69*(4), 839–872. https://doi.org/10.1111/lang.12353

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. https://doi.org/10.3758/s13428-012-0278-x

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. https://doi.org/10.3758/s13428-012-0242-9

Jiang, N. (2013). *Conducting reaction time research in second language studies*. Routledge. https://doi.org/10.4324/9780203146255

Jung, C. G. (Ed.). (1919). *Studies in word association: Experiments in the diagnosis of psychopathological conditions carried out at the psychiatric clinic of the University of Zurich*. https://doi.org/10.1037/13030-000

Kanellopoulou, C., Kermanidis, K. L., & Giannakoulopoulos, A. (2019). The dual-coding and multimedia learning theories: Film subtitles as a vocabulary teaching tool. *Education Sciences, 9*(3), 210. https://doi.org/10.3390/educsci9030210

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. https://doi.org/10.1080/09658211.2021.1924789

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14. https://doi.org/10.1037/a0021446

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143*(3), 1065–1081. https://doi.org/10.1037/a0035669

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

Landauer T. K., Foltz P. W., Laham D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284. https://doi.org/10.1080/01638539809545028

Lin, K. R., Wisman Weil, L., Thurm, A., Lord, C., & Luyster, R. J. (2022). Word imageability is associated with expressive vocabulary in children with autism spectrum disorder. *Autism & Developmental Language Impairments, 7*, 1–9. https://doi.org/10.1177/23969415221085827

Lindstromberg, S., & Eyckmans, J. (2021). The retrievability of L2 English multiword items in a context of strongly form-focused exposure: what matters? *Studies in Second Language Acquisition, 43*(5), 1040–1067. https://doi.org/10.1017/s0272263120000315

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*, 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559. https://doi.org/10.3758/BF03192726

Mirman D., Magnuson J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 65–79. https://doi.org/10.1037/0278-7393.34.1.65

Morett, L. M. (2019). The power of an image: Images, not glosses, enhance learning of concrete L2 words in beginning learners. *Journal of Psycholinguistic Research, 48*(3), 643–664. https://doi.org/10.1007/s10936-018-9623-2

Muraki, E. J., Abdalla, S., Brysbaert, M., & Pexman, P. M. (2022). Concreteness ratings for 62 thousand English multiword expressions. https://doi.org/10.31234/osf.io/m397u

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language, 31*(2), 237–260. https://doi.org/10.1016/0749-596x(92)90013-n

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. https://doi.org/10.3758/bf03195588

Nenadić, F., Podlubny, R. G., Schmidtke, D., Kelley, M. C., & Tucker, B. V. (2022). Semantic richness effects in isolated spoken word recognition: Evidence from massive auditory lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001208

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 4*(1), 32–38. https://doi.org/10.1016/s0022-5371(65)80064-0

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology, 76*. 1–25. https://doi.org/10.1037/h0025327

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255. doi:10.1037/h0084295

Patterson, A. (2021). Predicting second language listening functor comprehension probability with usage-based and embodiment approaches. *International Journal of Bilingualism, 25*(3), 772–788. https://doi.org/10.1177/13670069211000851

Patterson, A. (2022). The shape of a word: single word characteristics' effect on novice L2 listening comprehension. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2021-0157

Patterson, A. S., & Nicklin, C. (2023). L2 self-paced reading data collection across three contexts: In-person, online, and crowdsourcing. *Research Methods in Applied Linguistics, 2*(1). https://doi.org/10.1016/j.rmal.2023.100045

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*, 453–466. https://doi.org/10.3758/s13428-018-1171-z

Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass, 4*(3), 515–532. https://doi.org/10.1111/j.1747-9991.2009.00215.x

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279. https://doi.org/10.1017/s0140525x9900182x

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high-and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science, 31*(1), 157–168. https://doi.org/10.1080/03640210709336988

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition, 172*, 59–72. https://doi.org/10.1016/j.cognition.2017.12.007

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, *38*(4), 677–701. https://doi.org/10.1017/S0272263115000297

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*(3), 343–360. https://doi.org/10.1177/0267658310395851

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge University Press. https://doi.org/10.1017/9781108569057

See, A., Pappu, A., Saxena, R., Yerukola, A., & Manning, C. D. (2019). Do massively pretrained language models make better storytellers? *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 843–861. https://doi.org/10.48550/arXiv.1909.10705

Shaoul, C. & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413. https://doi.org/10.3758/brm.42.2.393

Shirai, Y. (2018). *Connectionism and second language acquisition*. Routledge. https://doi.org/10.4324/9780203118085

Sidhu, D. M., Kwan, R., Pexman, P. M., & Siakaluk, P. D. (2014). Effects of relative embodiment in lexical and semantic processing of verbs. *Acta Psychologica, 149*, 32–39. https://doi.org/10.1016/j.actpsy.2014.02.009

Smith, G. (2018). Step away from stepwise. *Journal of Big Data, 5*. https://doi.org/10.1186/s40537-018-0143-6

Spreen, O., & Schulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior, 5*(5), 459–468. https://doi.org/10.1016/s0022-5371(66)80061-0

Su, Y., Li, Y., & Li, H. (2022). Imageability ratings for 10,426 Chinese two-character words and their contribution to lexical processing. *Current Psychology*, 1–12. https://doi.org/10.1007/s12144-022-03404-4

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience, 26*, 858–866. https://doi.org/10.1101/2022.09.29.509744

Tillotson, S., Siakaluk, P., & Pexman, P. (2008). Body-object interaction rating for 1,618 monosyllabic nouns. *Behavior Research Methods, 40*(4), 1075–1078. https://doi.org/10.3758/BRM.40.4.107

Vitta, J. P., Nicklin, C., & Albright, S. W. (2023). Academic word difficulty and multidimensional lexical sophistication: An English-for-academic-purposes-focused conceptual replication of Hashimoto and Egbert (2019). *The Modern Language Journal*, *107*(1), 373–397. https://doi.org/10.1111/modl.12835

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. https://doi.org/10.3758/s13428-012-0314-x

Wise, R. J., Howard, D., Mummery, C. J., Fletcher, P., Leff, A., Büchel, C., & Scott, S. K. (2000). Noun imageability and the temporal lobes. *Neuropsychologia, 38*(7), 985–994. https://doi.org/10.1016/s0028-3932(99)00152-9

Wong, K. M., & Samudra, P. G. (2021). L2 vocabulary learning from educational media: Extending dual-coding theory to dual-language learners. *Computer Assisted Language Learning, 34*(8), 1182–1204. https://doi.org/10.1080/09588221.2019.1666150

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 72. https://doi.org/10.3389/fnhum.2012.00072

Zwaan, R. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*(1), 81–88. https://doi.org/10.1080/01638539909545070