Nickels, L., Lampe, L. F., Mason, C., & Hameau, S. (2022). Investigating the influence of semantic factors on word retrieval: Reservations, results and recommendations. *Cognitive Neuropsychology*, 39(3-4), 113-154.

Investigating the influence of semantic factors on word retrieval: reservations, results and recommendations

Lyndsey Nickels[1,2], Leonie F. Lampe[1,3], Catherine Mason[1,2], Solène Hameau[1,2]

[1] School of Psychological Sciences, Macquarie University, Sydney, Australia

[2] Macquarie Centre for Reading, Macquarie University, Sydney, Australia

[3] Department of Linguistics, University of Potsdam, Germany

Author Note

Lyndsey Nickels https://orcid.org/0000-0002-0311-3524

Leonie F. Lampe https://orcid.org/0000-0001-6020-8506

Catherine Mason https://orcid.org/0000-0002-6425-6262

Solène Hameau https://orcid.org/0000-0001-5282-2645

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence this research.

Solene Hameau was supported by an Australian Research Council Project grant (DP190101490) and Catherine Mason by a Macquarie University Research Excellence Scholarship (MQRES; scholarship number: 20191031).

Correspondence concerning this article should be addressed to Professor Lyndsey Nickels, School of Psychological Sciences, Macquarie University Sydney, NSW 2109, Australia. Email: lyndsey.nickels@mq.edu.au

Abstract

There is consensus that word retrieval starts with activation of semantic representations. However, in adults without language impairment, relatively little attention has been paid to the effects of the semantic attributes of to-be-retrieved words. This paper therefore addresses the question of which item-inherent semantic factors influence word retrieval. Specifically, it reviews the literature on a selection of these factors: imageability, concreteness, number of semantic features, typicality, intercorrelational density, featural distinctiveness, concept distinctiveness, animacy, semantic neighbourhood density, semantic similarity, operativity, valence and arousal. It highlights several methodological challenges in this field and has a focus on the insights from studies with people with aphasia where effects of these variables are more prevalent. The paper concludes that further research simultaneously examining the effects of different semantic factors that are likely to affect lexical co-activation, and the interaction of these variables would be fruitful, as would suitably scaled computational modelling of these effects in unimpaired language processing and in language impairment. Such research would enable refinement of theories of semantic processing and word production, and potentially have implications for diagnosis and treatment of semantic and lexical impairments.

Keywords: Semantic processing, lexical access, picture naming, aphasia

Speaking is an incredible skill. As Levelt (1992) noted, when we speak, we need to retrieve words at rates of 120–150 words/minute on average with spurts of up to 300 words a minute (Levelt, 1992, citing Maclay & Osgood, 1959, and Deese, 1984). This process is so efficient that we are unaware of the complexities involved in even retrieving a single word—only noticing when the right word does not come to mind sufficiently quickly. This paper focuses on this process—the retrieval of lexical items driven by conceptualisation of an idea. The paper is in two major parts, followed by a summary and conclusions section. In the first, we provide some important background, including highlighting the different methodologies used for investigating word retrieval, and the issues faced by each method, before turning attention to semantic effects on word production and the difficulties in measuring and defining semantic variables. The second part addresses the question of which item-inherent semantic properties influence lexical retrieval, with a focus on insights on this process from individuals with aphasia. We do not attempt a comprehensive, systematic review[1], but rather critically review the literature on those variables that have either been most studied and/or have been found to have the most impact on spoken word production.

Background Issues

Methods for investigating word retrieval

Much early research into lexical retrieval focused on the speech errors that arose when word retrieval was occasionally derailed in 'neurotypical' speakers (i.e., speakers with typical language skills and without language impairment; e.g., Butterworth, 1981; Dell, 1984; Fromkin, 1971, 1973b; Meringer & Mayer, 1895 (as cited in Levelt, 2001); Stemberger, 1982). However, naturally occurring errors are relatively rare: Levelt (2001) estimates no more than once or twice in every 1,000 words. They can also be subject to detection bias with some kinds of errors being easier to identify than others (see e.g., Cutler, 1981b; Fromkin, 1973a; Stemberger, 1992). Consequently, there is also a substantial body of work devoted to inducing errors, including tongue twister paradigms (e.g., Shattuck-Hufnagel, 1983, 1987; Stemberger & Treiman, 1986; Vitevitch, 2002), experiments inducing tip of the tongue states (e.g., Biedermann et al., 2008; A. S. Brown, 1991; R. Brown & McNeill, 1966), or requiring participants to respond faster than usual (e.g., Lampe, Hameau, & Nickels, 2022a; Moses et al., 2004; Vitkovitch & Humphreys, 1991). However, given that these methods to produce errors invoke processes that are not usual in spontaneous speech production, some authors, particularly within cognitive neuropsychology, have advocated and used an alternative approach capitalising on the more frequent errors that arise in language impairment (e.g., Dell et al., 1997; Kohn & Smith, 1995; Martin et al., 1989; Nickels, 1995). As we will discuss below, the factors that influence accuracy and error production in language impairment can provide further insights into lexical retrieval, but without the need for artificial experimental manipulations.

Researchers have also attempted to investigate the mechanisms underpinning lexical retrieval by using increasingly sophisticated methods of manipulating the context of word retrieval. These techniques aim to

[1] The decision not to provide a formal systematic review was motivated by several factors. First, that there were other papers (cited in the sections on individual variables) that already provided relatively systematic reviews of the literature on each variable. Second that systematic reviews may lead to the tendency for potentially important differences in methodology across studies to be obscured, and this is particularly problematic if a metanalysis is performed. Finally, a systematic review of all the relevant variables would have been too large in scope for a single article, not to mention the difficulty finding the appropriate articles with the relevant search terms without risking becoming either over, or under inclusive (e.g., 'semantic' and 'naming' as search terms in Scopus results in @14,000 hits).

influence processing at different levels of lexical retrieval and index this influence through effects on response latencies. These include, for example, priming of picture naming, picture-word interference and blocked cyclic naming methodologies. These tasks have sometimes required not only tortuously complex designs but equally complex inference from results to theory. Consider, for example, Levelt et al. (1991) where participants were provided with a picture to be named together with an auditory word which varied in its relationship to the target picture and in when this word was presented relative to picture presentation. Despite the process of interest being spoken word production, the dependent variable in this task was the latency of lexical decision to the auditory word (deciding whether the auditory stimulus was or was not a word). Thus, to interpret the results of this task, it is necessary to not only have a clear understanding of how lexical decision is hypothesised to occur, but also the nature of feedback and interaction within the language system, and particularly the relationship between auditory recognition and speech production processes. Consequently, any inference regarding processes in spoken word production on the basis of this task is heavily reliant on other assumptions about representation and processing.

Similarly, the picture-word interference task was developed as a variant of the classic STROOP task (where the ink colour of a written colour word has to be named, e.g., the word RED written in blue or red ink). In the picture-word interference task, the effects of differing relationships between a superimposed distractor word and a to-be-named picture (e.g., DOG or CAP written on a picture of a *cat*) have been used to investigate the factors influencing spoken word production, and in particular whether this process is competitive (e.g., Bürki et al., 2020; Glaser & Düngelhoff, 1984; La Heij, 1988; Lupker, 1979; Melinger & Abdel Rahman, 2013). However, once again, this task is complex and, in addition, there must be some mechanism to ensure the speaker names the picture rather than reads the word. Indeed, some authors argue that slower naming with a semantically related distractor does not reflect competition within the lexical system but instead attentional control mechanisms that ensure that only responses that are in the appropriate 'set' are produced (e.g., Mahon et al., 2007). Another task where non-linguistic factors have been argued to play a role is the blocked cyclic naming paradigm where pictures are named in sets of categorically related or categorically unrelated items over several repetitions (Belke et al., 2005). Relative to standard naming where each picture is named once in a random sequence, in blocked cyclic naming, working memory (e.g., Belke, 2008) and/or cognitive control (e.g., Abdel Rahman & Melinger, 2011; Belke & Stielow, 2013) have been suggested to play a larger role (see de Zubicaray et al., 2015, for further discussion of the different demands of standard and blocked cyclic naming).

Moreover, the more complex the task, the greater the possibility that a slight change in the task may alter the results (e.g., Damian & Martin, 1999; Nickels, 1997). Indeed, in some domains, a (cynical) observer may gain the impression that almost as much empirical work has been undertaken to better understand the tasks, as has used these tasks to gain insight into the processes they are believed to tap[2] (e.g., on Masked priming: Norris & Kinoshita, 2008; on Stroop: Hazeltine & Mordkoff, 2014; Kinoshita et al., 2017; Mills et al., 2019; on Stroop and picture-word interference: Starreveld & La Heij, 2017; on Lexical Decision: Norris, 2006). Shallice (1991) sums this up well:

"Empirical phenomena in the corresponding study of normal processes—human experimental

[2] We do not mean to imply that it is not important to understand how tasks work, merely to highlight that the simpler the task, the less research effort needs to be expended in working out the processes involved in that task, and the more research effort that can be spent using the task to develop theory.

psychology—are very slippery things. Many factors affect any experimental procedure. Make a slight change in one aspect—rate of presentation, stimulus material, recall delay, amount of practice, and so on—and the effect disappears or reappears, although according to theory, it should not. Therefore, even if a phenomenon is narrowly robust, the experimental result provides only a most insecure platform for theoretical inferences." (Shallice, 1991, p. 429)

Consequently, in this paper we focus on one of the least complex and most straightforward word production tasks—standard, simple, picture naming. In 'simple picture naming', a series of single pictures is presented for naming[3], with no distractors or primes nor any explicit manipulation of prior context and although participants are encouraged to name as quickly as they can, while maintaining accuracy, no particular time pressure is exerted.

In addition, rather than examining the influence of context on word retrieval, we instead focus on the critical variable approach[4] (Shallice, 1988), where the influence of stimulus properties (e.g., word frequency, word length) on word retrieval are used to inform our understanding of lexical processing. This approach was embraced in the literature given the difficulties with studying errors in neurotypical speakers, and with the development of technology that allowed relatively precise measurement of response latencies. There are now a number of studies (e.g., Alario et al., 2004; Valente et al., 2014) that have investigated the effects of numerous psycholinguistic variables using behavioural and electrophysiological measures aiming to understand exactly which sub-process of word production is influenced by what variable.

As the primary source of experimental evidence for conceptually-driven word retrieval is from object picture naming (with or without contextual manipulation), our focus will be on the retrieval of names for concrete objects. We briefly discuss broader issues relating to word retrieval for abstract concepts after our main discussions. We do not discuss verb retrieval, given the marked differences between nouns and verbs in their semantics, and refer the interested reader to the rich debate about differences between nouns and verbs and the interpretation of word class effects (e.g., Bi et al., 2005; Bird et al., 2000, 2003; Mätzig et al., 2009; Vigliocco et al., 2011). We do, however, note that at least some authors (e.g., Vigliocco et al., 2011) argue that nouns and verbs do *not* rely on different cognitive (or neural) systems.

Designs using the critical variable approach

Research taking the critical variable approach has used both factorial designs (e.g., Gardner, 1973; Holmes & Ellis, 2006; Howard et al., 1995; Mirman, 2011; Rossiter & Best, 2013) and/or examined statistical relationships using correlational or regression techniques (e.g., Carroll & White, 1973; Hameau et al., 2021; Lampe, Hameau, & Nickels, 2022b; Newcombe et al., 1965; Nickels & Howard, 1995; Oldfield & Wingfield, 1965). Factorial designs manipulate the variable of interest (e.g., frequency) while ensuring that 'nuisance' variables are held stable (matched) across the two groups (e.g., length). Comparison of naming latency and/or accuracy for sets

[3] Although this task could, accurately, be described as continuous picture naming, this term is often associated with the cumulative semantic interference literature where the number of previously presented items from a target's category is manipulated.

[4] Nozari and Pinet (2020) in their review of co-activation of representations describe this as the use of 'indices' to examine the effects of this co-activation.

of stimuli, that are, for example, high and low in the manipulated variable, allows determination of whether this variable has an effect on performance.

However, Anne Cutler (1981a) commented on the limitations of the factorial approach in her beautifully entitled paper “Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990?”. She highlights the difficulty of manipulating one variable while matching others, and indeed, even when this is possible, the resulting stimuli may be “odd”—unrepresentative of the stimuli as a whole (Ellis et al., 1996). For example, high frequency words are usually also acquired early in life. Consequently, if attempting to manipulate both variables, some cells are therefore populated with unusual items: For example, stimuli that are low in age of acquisition and low in frequency, may be overly represented with words from fairy tales (e.g., witch, fairy). Moreover, the dichotomisation of continuous variables, that has to occur in factorial designs, has also been associated with a loss of statistical power (e.g., Baayen, 2004; Cohen, 1983). Cutler (1981a) also noted that more and more potential confounding variables continued to be discovered across the 1970s, a pattern certainly that has continued over the subsequent 40 plus years. This has resulted in increasing difficulty in the matching process.

As an alternative to factorial designs, statistical methods examining the extent to which variables or combinations of variables predict performance have also been used. In such regression approaches, to-be-controlled variables are included as covariates in the statistical models, such that variability caused by these variables is controlled when assessing effects of variables of interest (Baayen, 2004). These types of analyses have been particularly influenced by the availability of increasingly sophisticated regression techniques and facilitated by the exponential increase in (accessible) computing power to perform these regressions.

Probably the first example of this technique was Oldfield and Wingfield (1964, 1965) influential report of the effect of frequency on picture naming latency in neurotypical participants: Correlations showed a strong negative linear relationship between naming latency and log frequency. As an aside, despite its impact on the field, it is interesting to note that this study asked just 11 participants to name only 26 items, a study that would now be considered hugely underpowered with a potentially unrepresentative item sample. Importantly, Oldfield and Wingfield were aware that there was a confound between frequency and word length: Compare, for example, their three highest frequency words, book, chair, and shoe, and the three lowest frequency words, metronome, gyroscope, and xylophone. Consequently, they carried out partial correlations, examining correlations between frequency and naming latency when either number of syllables or number of letters were held constant (the correlations remained significant), or correlations between word length (either syllables or letters) and naming latency when controlling for frequency (no significant correlations).

However, as already noted, a growing number of variables have now been argued to affect spoken word production, and the results of Oldfield and Wingfield (1964, 1965) seminal work increasingly called into question (e.g., Carroll & White, 1973; Lachman et al., 1974). For example, Carroll and White, using multiple regression, argued that the effects of frequency were in fact ‘really’ effects of age of acquisition. However, given that the correlations between their measures of frequency and age of acquisition ranged between r = .645 and r = .748, there are potential concerns regarding problematic effects of multicollinearity in Carroll and White’s analyses. We will return to this issue below.

Johnson et al. (1996) also exhorted that “research on item attributes and participant characteristics would also benefit from more robust multivariate approaches sensitive to the undoubtedly complex intercorrelations among these variables.” (Johnson et al., 1996, p. 130). Indeed, there have now been examinations of the effects of multiple psycholinguistic variables and across many languages (see Perret & Bonin, 2019, for a meta-analysis of 14 such studies).

However, regression techniques are certainly not without problems. For example, Ellis et al. (1996) illustrate how regressions run on a single administration of a set of items for naming in people with aphasia may produce unreliable results—reinforcing the importance of replication of effects within an individual. They also note that multicollinearity likely played a role in the lack of stability. When variables are highly correlated, small changes in exactly which items are named correctly and which incorrectly can result in marked changes in the pattern of significant variables (see also Morrison, 2003).

It is well established that very high correlations between independent variables (i.e., $r \geq .9$) can be problematic, in that they can lead to instability in statistical measurement (large fluctuations of the regression estimates and standard errors). But caution is needed even for lower correlation coefficients (e.g., $r \geq .7$ between two independent variables is not recommended; Tabachnick & Fidell, 2019) and when more than two predictors in a model are correlated as weakly as .25 (e.g., Vatcheva et al., 2016). If variables are intercorrelated, their effects are confounded, and, as a result, assessment of their individual importance is difficult to estimate. As such, the outcomes of models with such multicollinearity can be unreliable and inferences about the statistical significance of highly correlated variables can be incorrect (see Morrison, 2003, for an example and discussion of this in practice). Note that how high the correlation between predictors has to be before variances are inflated depends on sample size, with larger samples (number of stimuli and participants) promoting lower levels of multicollinearity. However, given that there is no rule of thumb regarding how high the correlation needs to be to cause a concern regarding multicollinearity, correlations between predictors should not be taken as the sole indicator of multicollinearity.

Variance Inflation Factors (VIFs) are a way to estimate potentially harmful levels of multicollinearity between variables and to identify instances where the combination of different variables in a statistical model predicts another variable in the model (Hutcheson & Sofroniou, 1999). VIFs directly measure the extent to which the variance of a given coefficient (its standard error) is inflated due to multicollinearity. VIFs can be calculated for regression models and values above a certain cut-off (e.g., 2.5, Allison, 2012; 5.0, Hair et al., 2014; Rogerson, 2011) indicate compromising effects of multicollinearity.

When there are many correlated independent variables, another approach to solve the associated problems can be to use principal component analysis to reduce these variables into a smaller number of factors, each correlating with a number of the initial correlated variables (e.g., Butler et al., 2014; Halai et al., 2017; Mirman et al., 2015). This approach makes it possible to obtain orthogonal factors that represent the original independent variables, therefore solving the multicollinearity issue. However, interpretation of a given obtained factor depends on how much the combination of variables behind this factor "makes sense" (Tabachnick & Fidell, 2019), that is, how much they are effectively correlated to each other (and relatively uncorrelated to the rest of the variables and factors), and how much their combination is theoretically relevant. In addition, this procedure does not allow one to interpret the effect of one single variable within a factor.

Not all picture naming is equal!

There are many papers now published using a simple picture naming paradigm where context is not manipulated, however, the careful reader will have realised that even within this literature there are methodological differences that have the potential to influence results.

Familiarisation & Repetition.

In the neurotypical literature, the most common methodological difference across studies is whether or not familiarisation is used. Familiarisation usually comprises showing the participants the pictures and informing them of the target name. Coming from a clinical background working with people with aphasia, the idea that someone would 'train' the required name prior to testing seems bizarre! Indeed, it is rare to find studies with people with aphasia that use this paradigm. Yet this is surprisingly common in the neurotypical literature. Indeed, Alario et al. (2004) explicitly motivate the use of familiarisation in their study of the factors affecting picture naming on the basis that most psycholinguistic experiments involving picture naming use a familiarisation process.

The idea behind the use of familiarisation is that researchers are interested in the time course of spoken word production from a conceptual representation. Consequently, familiarisation overcomes variability in naming latencies due to poor pictorial representations of the target and slow recognition and ensures that the correct concept is accessed. However, intuitively, the reaction time to a picture where the participant has to overcome their instinctive reaction to remember the assigned name seems unlikely to be representative of the 'usual' naming process. Consider, for example, a picture of a couch/settee/sofa—where the target is 'couch'—the participant who naturally would call this 'sofa' now has to inhibit this name and recall the correct name that they have been told.

There is also variability in what 'counts' as familiarisation. For example, Alario et al. (2004) used naming followed by the correct name being provided and others include a study period where the pictures are presented together with the target word but without naming being required (e.g., Rabovsky et al., 2016), while some studies (e.g., Lampe, Hameau, & Nickels, 2022a; Llorens et al., 2014; Rabovsky et al., 2021) have a first round of naming without correction. Both types of familiarisation, as Alario et al. note, would be expected to result in faster naming times, a claim supported by their data (see also Lampe, Hameau, & Nickels, 2022a, Appendix B).

In terms of the effects of familiarisation on the factors affecting naming, Alario et al. (2004) emphasise the similarities between the results of their first (familiarisation with correction) round, and second naming round. Nevertheless, they did find significant effects of familiarity in the first naming round, but not in the second, and of visual complexity and (reverse effects) of number of syllables in the second but not in the first. Hence, it appears that there may be differences in the effects seen between the first round of naming and the second, after correction of incorrect naming attempts. However, of course, in this study the effects of repetition and correction are confounded.

In a study that enables the effects of repetition to be addressed, Lampe, Hameau, and Nickels (2022a; Appendix B) compared the effects of psycholinguistic variables in two rounds of standard picture naming. Unsurprisingly, responses were faster and more accurate in the second compared to the first standard naming round. Effects of psycholinguistic variables on naming accuracy were unaffected by the repetition, which resembled the non-significant interactions between number of semantic features or intercorrelational density and task repetition in Rabovsky et al. (2021). However, in the naming latency analysis, some effects were attenuated in the second compared to the first naming round. There was a significant interaction between repetition and number of semantic features with an attenuated effect of this variable in the second standard naming round compared to the first exposure to the items. The magnitude of the effects of the other semantic variables investigated did not differ between the two rounds of standard naming. In addition, the effects of some of the control variables (i.e., name agreement, image agreement, frequency, familiarity, and ordinal category position) were somewhat weaker in the second standard naming round (Lampe, Hameau, & Nickels, 2022a; Table B.1). Importantly, all effects in the second standard naming round were consistent with the first naming round, although some were weaker, and there was no reversal of the direction of effects in the second standard naming round. However, caution may be in order, as Rabovsky et al. (2021), in contrast to Lampe, Hameau, and Nickels (2022a), found *increased* effects of a semantic variable (intercorrelational density) in a second naming round.

In sum, effects of familiarisation and/or repetition can affect the results found when examining how psycholinguistic variables influence picture naming, but we cannot yet be confident in what ways. At the very least, it may lead to underestimation of the effects.

Speeded naming.

In the literature on picture naming in neurotypical adults, some studies have also used a speeded version of the task that requires participants to prioritise naming speed over accuracy. Specifically, paradigms have been used where an individual either has to attempt to respond before a deadline (e.g., 500 or 600ms after picture onset; e.g., Damian & Dumay, 2007; Lampe, Hameau, & Nickels, 2022a; Lloyd-Jones & Nettlemill, 2007; Moses et al., 2004; Vitkovitch et al., 1993; Vitkovitch & Humphreys, 1991), or respond at a particular (fast) tempo set by a series of beeps (e.g., Fieder et al., 2019; Mirman, 2011; Mirman, Kittredge, et al., 2010). It is of note that many of the studies examining semantic variables that have shown significant effects in neurotypical adults tend to have been those that have used speeded naming (for discussion see Lampe, Hameau, & Nickels, 2022a). Moreover, speeded naming has been suggested to cause enhanced effects of item-inherent word characteristics due to disruptions of cognitive control and resulting modulations of responsiveness to input (e.g., Mirman, 2011).

Lampe, Hameau, and Nickels (2022a) directly compared effects of psycholinguistic variables in speeded and standard picture naming. The majority of control variables and some of the semantic variables investigated showed an interaction between task and the effects of that variable (name agreement, age of acquisition, image agreement, frequency, familiarity, imageability, distinctiveness)[5]. Importantly, these interactions were due to differences in the strength of the effects rather than indicating a change in the direction of effects. Consequently, it may be that speeded naming could be a valid method by which to amplify the effects of psycholinguistic effects on naming in neurotypical participants.

Cumulative semantic interference.

It is now well established that when naming a series of pictures, having previously named one item from a category will slow the naming of other items from the same category and this is cumulative with the more items that are presented (e.g., Howard et al., 2006). Given the limited number of pictureable items, most naming sets will have several items from the same category and consequently cumulative semantic interference will be a confound. This is something that does not seem to have been addressed in the literature to date—presumably

[5] The effects were stronger on latencies in the speeded naming task than in the standard naming task for name agreement (facilitatory in both tasks), age of acquisition (inhibitory only in speeded naming), image agreement (facilitatory in both tasks), frequency (facilitatory only in speeded naming), familiarity (facilitatory in both tasks), trial order (facilitatory only in speeded naming), and distinctiveness (inhibitory in both tasks). Although five more semantic variables (number of semantic features, intercorrelational density, number of near semantic neighbours, semantic similarity, typicality) did not show a difference between tasks, they also did not show significant simple effects on either task.

because it is hard! Unfortunately, there are simply not enough categories to obtain a sufficiently large set of items that are from different semantic categories (and/or are not semantic neighbours): Lampe, Hameau, and Nickels (2022b) could only obtain a set of 35 items using this method. However, they did attempt to control for the effects by including as a covariate in the analysis a count of the number of previous items that had been presented within the category (ordinal category position) and we would suggest that this may be a fruitful way forward for other authors.

Deadlines and Coding.

In experiments with people with aphasia, there are several choices that can be made with marked variability across (and within) research groups. While there is no evidence that we are currently aware of to suggest that these decisions necessarily affect the factors found to affect picture naming responses, equally we cannot assume that they do not. First, analogous to the difference between speeded and standard naming in neurotypical participants, is the choice of deadline for naming which ranges from 5 seconds (e.g., Lyalka et al., 2020) to no time limit (e.g., Nickels & Howard, 1994, 1995). This has recently been investigated by Evans et al. (2020) who, looking at picture naming response time distributions of people with aphasia estimated optimal cut offs of around 5 to 10 seconds. They suggested that cut offs of shorter than 10 seconds could lead to (some people) with aphasia being under speeded naming conditions, with the potential issues discussed above. Another methodological choice relates to whether it is the first or the final response that is coded, and what is to be considered a first response. Once again this varies, but to our knowledge the effect on results has not been systematically investigated.

To conclude, while these methodological differences are likely to persist in the literature on picture naming, what is essential is that researchers not only make them explicit in their reporting, but also those evaluating and reviewing the literature are careful to take them into account. Otherwise, there is the danger that potential variability caused by differences in methodology may be interpreted as reflecting differences in cognitive processing.

Semantic effects on word production

Much of the previous research on psycholinguistic factors influencing word retrieval has focused on the influence of lexical and phonological factors, and it is of note that despite the centrality of semantic processing to word production only one unambiguously semantic variable was included in Perret and Bonin's (2019) meta-analysis of factors affecting spoken picture naming: imageability[6]. The relative lack of attention in the recent literature to item-inherent semantic variables on picture naming is in stark contrast to the enormous research effort, described above, on the influence of semantic context on word production. Hence, the focus of this paper is on the effects of these semantic variables on simple picture naming.

One possible reason for the lack of attention is that the effects of these variables appear to be small and elusive in word production in neurotypical speakers (for discussion see Lampe, Hameau, & Nickels, 2022b). For example, effects of semantic neighbourhood density have only been found when participants have to name

[6] Perret and Bonin (2019) do also locate conceptual familiarity at this level of processing. Note too that in their analysis they combine effects of imageability and image variability as a single variable (see section on imageability effects for further discussion).

before a deadline or at a fast tempo in speeded picture naming (e.g., Fieder et al., 2019; Mirman, 2011) but not in simple picture naming (e.g., Bormann, 2011; Hameau, Nickels, et al., 2019; Lampe, Hameau, & Nickels, 2022b). Interestingly, in neurotypical speakers, compared to the relative paucity of research on spoken word production, there has been more focus on the effects of semantic variables on word recognition (e.g., lexical decision), see, for example, the work by Penny Pexman, Melvin Yap and colleagues (e.g., Pexman, 2012, 2020; Yap et al., 2012; Yap & Seow, 2014). Given the very different demands of word recognition and production, we do not focus on this literature here. Compared to the limited effects of semantic variables on picture naming in neurotypical speakers, where they have been studied, these effects seem to be more commonly found in people with aphasia. For example, in Hameau, Nickels, et al.'s (2019) study of semantic neighbourhood, no significant effects were found in unimpaired participants, but effects were apparent in people with aphasia. Consequently, in the remainder of this paper, we focus on evidence for effects of semantic variables on word retrieval in people with aphasia. Before we turn to this evidence, we first will consider the challenges of interpreting these effects in terms of theories of spoken word production.

Challenges in theoretical interpretation

Semantic processing and representation is complex and is possibly the area within theories of spoken word production where there remains the least consensus[7]. For example, there is debate regarding whether there is a distinction between lexical and conceptual semantics (e.g., Johnson et al., 1996), and the nature of semantic representations themselves. Moreover, there is huge debate in the literature regarding the nature of semantic organisation itself. While we have a focus here on word production rather than semantic processing per se, and a review of this literature is beyond the scope of this paper, we will briefly address semantic organisation given that representation at this level has knock on effects to word retrieval.

In the context of language impairment, probably the most influential cases in the debate have been (semantic) category-specific disorders, with two major accounts of semantic organisation—categorical and (de)compositional. Under the categorical account, different categories of concept (e.g., proper vs common nouns, animate vs inanimate objects, abstract vs concrete nouns) are claimed to be independently represented (in mind and/or brain). Consequently, these categories can be selectively impaired and double dissociations are possible. Independent representation and/or processing of semantic categories can occur at different levels (e.g., semantic, lexical), accounting for different patterns of category specific impairment.

In contrast, what we term (de)compositional accounts propose that semantic representations of words are comprised of networks of components of meaning, such as semantic features or links between related concepts. These components may be of different types, for example relating to visual or other sensorimotor attributes of an item, functional properties or (contextual) associations. The (de)compositional accounts include those where undifferentiated features are distributed in semantic space (e.g., Dell et al., 1997; Plaut & Shallice, 1993; Vigliocco et al., 2004). Other decompositional accounts propose that different sensory modalities of semantic knowledge are stored in separate but linked modules (e.g., Allport, 1985) that may include an amodal

[7] Much work on semantic representations and semantic processing conducted outside the area of word production (e.g., Cree et al., 1999; Farah & McClelland, 1991; Masson, 1995; Plaut & Shallice, 1993) has not (yet) been incorporated in theories of word production, perhaps because most experimental findings in word production research can be explained in the context of highly simplified semantic representations (Vinson et al., 2013).

'hub' (e.g., Patterson et al., 2007). Lastly, there are 'embodied' or 'grounded' models where semantics is rooted in simulation of sensory and motor properties of a concept (e.g., Barsalou, 2008, 2016, 2020; Kiefer & Pulvermüller, 2012; see Meteyard et al., 2012, for a review).

Of course, the categorical and (de)compositional accounts are not completely dissociable. Indeed, many authors, who at first glance might seem to be advocating a categorical approach, in fact argue that the difference in the semantic characteristics of the categories is what underpins this organisation (e.g., Warrington & Shallice, 1984). Furthermore, this distinction does not capture the full range of proposals regarding semantic organisation (for reviews of different accounts see e.g., Vigliocco & Vinson, 2007; Vinson et al., 2013). Other accounts include unitary (non-decomposed) but non-hierarchical semantic representation (e.g., Collins & Loftus, 1975; Levelt et al., 1999) or distributional accounts where meaning is represented by the statistical distribution of the contexts across which words appear (e.g., Landauer & Dumais, 1997; Lund & Burgess, 1996).

For the purposes of our paper, with its focus on lexical retrieval in spoken word production, the key point is that, whatever the precise theory of semantic processing, activation of a concept results in co-activation of other concepts that are related in meaning (be that categorically related, e.g., desk—table, or contextually/thematically related, e.g., desk—school). Similarly, there is broad consensus that semantics influences lexical processing—it is generally agreed that rather than one concept being selected and this concept alone activating the lexical system, there is co-activation of semantically similar lexical items (e.g., Dell et al., 1997; Hillis & Caramazza, 1995; Levelt et al., 1999). Moreover, researchers concur that lexical activation and/or selection is influenced by the semantic properties of an item to some degree—whether this is, for example, by the 'richness' of the semantic representation, the other semantic representations that are co-activated or the similarity between co-activated semantic representations.

In addition to the considerations above, there is also the fact that when considering the effects of item-inherent variables on spoken word production, it is common to equate effects to particular levels of processing (e.g., Alario et al., 2004). The variables we are focusing on here are, by definition, related to semantic representation and processing. The source of their effect is therefore at the level of semantics. However, it need not be the semantic level at which the variable has its influence on word production, and the level at which these effects 'play out' in terms of influencing word production speed and accuracy will depend on the theory. For example, a variable that results in stronger semantic activation will result in stronger activation of the lexical form, which in a cascading model, could, in turn, result in stronger activation of that concept's phonemes. Hence this semantic variable would have an influence at all three levels. In addition, if selection occurs at the lexical level, then those items with stronger activation from semantics, by virtue of higher values in semantic factors, will be facilitated. In other words, there is a semantic source but lexical origin of this effect.

Similarly, with regards to the level at which variables have their effects on spoken word production in people with aphasia, it is common to equate effects of semantic variables (such as imageability) with semantic impairments. However, once again, the picture is not so straightforward. It is true that an individual with a semantic impairment, under most accounts, would be predicted to show better performance on items which are higher in 'helpful' semantic variables (e.g., higher in imageability). However, it is also the case that an individual with difficulty activating lexical forms would also show better performance on these same items. Similarly, semantically related errors are thought to arise from lexical as well as semantic impairment (Caramazza & Hillis, 1990). Consequently, while we are interested in whether semantic variables affect spoken word production in people with aphasia, these variables could have their effects at a number of different levels, in individuals with different impairments. Consequently, we do not focus on this aspect of the data to any great extent, given our interest on what this literature can tell us about semantic representation.

Quantifying facets of meaning

It is likely that another contributor to the infancy of the research on semantic influences on lexical retrieval in spoken word production is the difficulty in measuring the relevant factors. Some psycholinguistic variables can easily be counted (e.g., number of phonemes, frequency of occurrence) and others can be calculated (e.g., orthographic neighbourhood density). However, this is not true of many semantic variables which require the relationship between items to be determined. For example, how many semantic neighbours does an item have, how close are these neighbours, how similar are the properties of an item to those of another item, or to all the others in our semantic system? In addition, as with most psycholinguistic variables, determining the appropriate measure requires a theoretical position to be taken. For example, is the relevant measure type or token frequency, the frequency of the stem or of the surface form? Nevertheless, when deciding which semantic factors may be relevant, relative to other psycholinguistic variables, the theoretical domain is, most probably, more complex with a larger number of dimensions and, arguably, less consensus (e.g., are semantic representations comprised of semantic features?).

Some authors have used semantic features generated by participants (from e.g., Devereux et al., 2014; McRae et al., 2005; Rosch & Mervis, 1975; Vinson & Vigliocco, 2008) to determine semantic relationships. Others have measured semantic relationships in terms of the co-occurrence within corpora (e.g., Latent Semantic Analysis (LSA), Landauer et al., 1998; Continuous bag-of-words model (CBOW), Mikolov et al., 2013; Global Vectors for word representation (GloVe), Pennington et al., 2014; see Günther et al., 2019, for review), or associations (e.g., de Deyne et al., 2019; Nelson et al., 2004), or have used ratings (e.g., Bormann, 2011; Rossiter & Best, 2013; Warriner et al., 2013). While there may be a temptation to assume that 'objective' measures (e.g., from co-occurrence corpora) are 'better', it is important to realise that any measure is only an approximation. Indeed, it is well known that, in frequency measures, personal experience influences the 'true' frequency of an item for any participant. This is also true for semantics, where it will influence the richness of the semantics of an item and its associations. For example, in ratings of valence the words 'cat' or 'religion' could be strongly positive, negative or neutral depending on your view. Critically, this is more than an academic issue, as it influences how we respond. For example, one man with aphasia responded 'blue' to a picture of a dog—influenced by the name of his own dog; another man produced the low frequency word 'Leica' for a picture of a camera—again, most probably, influenced by his personal experience, and consequently higher than normal subjective frequency of the word. Hence, all measures are noisy and one should not automatically assume that measures that seem more objective are necessarily less noisy than those, for example, from ratings, or participant-generation. Indeed, there is evidence that in some cases participant rated measures may better account for variance in performance than objective counts. For example, both Gordon (1985) and Gernsbacher (1984) found that, for low frequency words, subjective ratings of a word's frequency (familiarity) were better predictors of lexical decision latency than objective frequency counts.

For many semantic variables, the fact that they may be derived in different ways is often overlooked and the different operationalisations used interchangeably. However, how valid it is to equate different operationalisations, is open to debate. For example, Hameau, Nickels, et al. (2019) looked at the issue of semantic neighbours, defined as the number of words of similar meaning in the lexicon. They found that in the literature six different ways of measuring neighbourhood for a given word had been used. The resulting values were, perhaps surprisingly, at most moderately correlated (correlation coefficients ranging from r = -.112 to r = .376). Table 1 provides an example of the neighbours that are derived from the different measures (see also Table 2 in Hameau, Nickels, et al., 2019). It is clear that the different measures provide not only different absolute numbers of semantic neighbours, but also the neighbours themselves differ, with many not fitting with (at least our) intuitions of what is related.

In this review, as noted above, we do not attempt a comprehensive survey of all semantic variables, but focus on a selection that have received closer attention and/or proved more influential for spoken word production and could be considered 'item-inherent' in so far as they are properties of the lexical items themselves. These are summarised in Table 2.

In the sections that follow, we first define the variable, and briefly note whether the effects of this measure have been studied in picture naming in neurotypical participants. This is followed by the discussion of these effects in naming in people with aphasia. We focus on picture naming (or other tasks involving word production from conceptual processing) but where the data is limited in picture naming, we do note the patterns in other tasks. Similarly, if the evidence is restricted in people with aphasia, we note evidence from other populations (e.g., Alzheimer's disease). Finally, in each section we discuss the suggested explanations for these findings.

Table 1

Neighbours of the words "owl" and "bench" across different measures

	owl		bench	
	number	neighbours list	number	neighbours list
Rated competitors[1]	5.74	NA	4.04	NA
Feature-based near neighbours[2]	24	(including) falcon, partridge, nightingale, hawk, pigeon, oriole, blackbird, sparrow, raven, bluejay, starling, goose, dove, pheasant…	6	Bat (baseball), board (wood), chair, peg, stick, stool (furniture)
Feature-based distant neighbours[2]	108	(including) alligator, barrel, beans, bear, beaver, beetle, bison, board, boots, box, bread, buffalo, bull, butterfly…	294	(including) airplane, alligator, anchor, apron, armour, asparagus, bag, ball, balloon, banana, banjo, banner, barn, basket…
Neighbours determined from context (Continuous Bag of Words)[3]	70	(including) frogmouth, tawny, owlet, bubo, jackdaw, bird, limpkin, nightjar, mynah, myna, catbird, woodpecker, waxwing, lemur…	7	sitting, aldermanic, chair, sit, table, sat, touchline.
Neighbours determined from context (Latent Semantic Analysis)[4]	19	mouse, owls, hooted, bray, meow, mice, donkey, cat, griggs, paws, cosgrove, lion, bathrobe, nibble, aha, beezus, quimby, ramona, whiskers	9	frits, sal, cass, alf, golly, superkids, tic, benches, bus
Neighbours from association norms[5]	8	bird, hoot, night, eyes, wise, animal, barn, eye	14	seat, park, sit, chair, mark, wood, press, work, beach, sitting, stool, table, warmer, wooden

[1] Ratings of number of competitors on a scale of 1 to 7: e.g., Bormann (2011).
[2] Near neighbours are words with cosine similarity >= .4; distant neighbours are words with cosine similarity between 0 and 0.25 in semantic feature norms (McRae et al., 2005): e.g., Lampe, Hameau, and Nickels (2022b), Mirman (2011).
[3] Neighbours determined by as words with less than .6 cosine distance (ordered by increasing distance) in co-occurrence patterns in text corpora (Mandera et al., 2017; see also http://meshugga.ugent.be/snaut//)
[4] Near neighbours determined from text corpora: http://lsa.colorado.edu/ (default settings; words with at least .4 similarity with the target).
[5] Neighbours determined by responses given by at least 2 participants in semantic association norms (SWOW, (de Deyne et al., 2019; see also https://smallworldofwords.org/en/project/research) in response to the target word as a cue.

Table 2
Overview of item-inherent semantic variables reviewed in this paper.

Variable	Definition	Examples of studies of effects on spoken word production in aphasia
Imageability	The extent to which a word/concept is associated with a mental image.	Nickels & Howard, 1995 Nickels, 1995 Howard & Gatehouse, 2006
Concreteness	The degree to which a word/concept is associated with a physical referent.	Nickels & Howard, 1995 Nickels, 1995 Howard & Gatehouse, 2006 Franklin et al., 1995
Number of semantic features	The number of semantic features/information associated with a word/concept.	Lampe et al., 2021
Typicality	The extent to which a word/concept is a "typical" representation of its semantic category, e.g., 'sparrow' is a highly typical representation of a "bird" while 'penguin' is an atypical representation of the category "bird"	Laiacona et al., 2001 Rossiter & Best, 2013 Rogers et al., 2015 Lampe et al., 2021
Intercorrelational density	Extent to which feature pairs of a word/concept co-occur together across other words/concepts. Words/concepts with higher intercorrelational density have features that commonly co-occur across words/concepts, e.g., the features *swims* and *has fins or swims* and *has scales* of the concept 'goldfish' would frequently co-occur together	Lampe et al., 2021
Featural distinctiveness	A measure of how distinctive a feature is; highly distinctive features are shared by few words/concepts while low distinctives features are shared by many words/concepts.	Mason-Baughman, 2009 Wallace & Mason-Baughman, 2012
Concept distinctiveness	Mean distinctiveness of all of a word's/concept's features. Higher concept distinctiveness means a word/concept has more unique features, lower concept distinctiveness means a word/concept has fewer unique features.	Lampe et al., 2021
Animacy	The extent to which a word/concept is associated with being alive and/or sentient.	Caramazza & Hillis, 1991 Laiacona & Capitani, 2001 Barbarotto et al., 1995
Semantic neighbourhood density	Number of (near) semantic neighbours of a word/concept, variably captured through associates, shared features, or category co-ordinates.	Lampe et al., 2021 Hameau, Nickels, et al., 2019 Hameau, Biedermann, et al., 2019

		Mirman & Graziano, 2013 Mirman & Magnuson, 2008 Mirman, 2011 Lampe et al., 2017 Bormann, 2011 Bormann et al., 2008 Kittredge et al., 2007a, 2007b Blanken et al., 2002
Semantic similarity	The mean featural similarity between a target word/concept and other words/concepts in the mental lexicon. Words/concepts with higher levels of semantic similarity share their features on average with a higher number of other words/concepts than those with lower semantic similarity.	Lampe et al., 2021
Operativity	The degree to which a word/concept is able to be touched, manipulated or experienced by the senses.	Gardner, 1973 Nickels & Howard, 1995
Valence	The degree to which a word/concept is associated with positivity or negativity (also sometimes defined as pleasantness/unpleasantness)	
Arousal	The degree to which a word/concept is associated with activity or passivity	

Imageability & Concreteness

One of the earliest semantic factors investigated was *imageability*. It came to the fore as a result of Pavio's influential work on memory (e.g., Paivio, 1971), where words with high imageability values were found to be better remembered and recognized than those with lower imageability (e.g., Paivio, 1969). Imageability is usually rated on the basis of how easy it is to create a visual or auditory image of the referent corresponding to the word (e.g., Medical Research Council (MRC) database, Coltheart, 1981; Bristol norms, Stadthagen-Gonzalez & Davis, 2006). Some authors, particularly working in French, have used a different rating—*image variability*, which was originally investigated by Snodgrass and Vanderwart (1980) (e.g., French norms from Alario & Ferrand, 1999; Bonin, Peereman, et al., 2003, used in Alario et al., 2004; Bonin et al., 2002; but see Bonin et al., 2011, 2020; Bonin, Méot, et al., 2003).

In image variability, the rating is of whether a given word evokes few or many different images for that particular object. However, Bonin et al. (2011) found that image variability and imageability ratings for the same items only correlated .43, and Nishimoto et al. (2012) in Japanese found even lower correlations between the two (r = -.02). This is in contrast to different ratings of imageability which correlated between .75 and .93 (Bonin et al., 2011). Consequently, these two measures are likely to tap different aspects of processing and it is unwise to consider them equivalent. Unfortunately, this was the case in both of the two influential examinations of effects of psycholinguistic variables on naming in unimpaired participants: Alario et al. (2004) used only image variability, but called it imageability and equated the two measures; Perret and Bonin (2019) in their Bayesian metanalysis combined studies that used imageability (e.g., Ellis & Morrison, 1998) and those that used image variability (e.g., Bonin et al., 2002; Bonin, Peereman, et al., 2003; Perret & Laganaro, 2013). Nevertheless, there seems to be evidence that both imageability (e.g., Barry et al., 1997; Ellis & Morrison, 1998) and image variability influence picture naming speed in neurotypical participants: Items rated as having higher imageability or higher image variability are named faster.

Concreteness is a closely related variable, rated on the basis of how available the referent of the word is to sensory experience. This has been less widely investigated in picture naming in neurotypical adults (although there is a large literature on effects on learning and lexical decision), and indeed Perret and Bonin (2019) did not include (nor mention) concreteness in their meta-analysis. This dearth of research is most likely because the early research on this variable viewed it as a categorical (abstract vs concrete) rather than a continuous measure, and of course pictureable items are by definition concrete. There has however been research which has found effects of concreteness on, for example, tip of the tongue effects (e.g., Gianico-Relyea & Altarriba, 2012).

Concreteness and imageability are strongly correlated (e.g., Benjafield & Muckenheim, 1989; Bird et al., 2001; Bonin, Méot, et al., 2003; Toglia & Battig, 1978). For example, Rofes et al. (2018, p. 118) cite a correlation of .84 across all the items in the MRC psycholinguistic database (Coltheart, 1981), although the correlations are often lower in sets of pictureable items (most likely due to the more restricted ranges: e.g., Nickels and Howard (1995) $r = .57$ (set 1) and .46 (set 2)). In the past, many authors have used concreteness and imageability interchangeably, have not distinguished between these two variables and/or assumed that they map onto the same underlying construct (Kousta et al., 2011). Nevertheless, in Paivio's dual coding theory (Paivio, 1971, 1986, 1991, 2007) a distinction is made between the two. In this theory, words have two representations: a verbal/language-based "logogen" representation and a non-verbal/imagery-based "imagen" representation. Words with high imageability are said to have both a logogen and imagen representation while low imageability words are suggested to have only a logogen representation (or a weak imagen). Crucially, Paivio et al. (1968) noted two exceptions to the generally high correlation between concreteness and imageability. The first were words with high concreteness but low familiarity which tended to have low imageability (most likely due to individuals knowing that something is an object but being unaware of what the object looks like; e.g., armadillo, quokka). The second were words that had emotional connotations, which despite having very low concreteness ratings, tended to have unusually high imageability ratings. In addition, Kousta et al. (2011) found that concreteness and imageability had different distributions with concreteness showing a bimodal distribution (items either rated as abstract or concrete) and imageability unimodal (items show a graded range of ratings) (see also Paivio et al., 1968).

Research is relatively limited on effects of imageability and/or concreteness on word retrieval in aphasia, despite a relatively large literature on their effects on reading and comprehension (see for example the work by Warrington, Crutch and colleagues, e.g., Crutch, 2006; Crutch & Warrington, 2007, 2010; Crutch et al, 2006; Warrington, 1981; Warrington & Shallice, 1984) which can largely be attributed to the emphasis on picture naming as the primary methodology. Abstract words are harder to elicit, particularly in the absence of other language cues (e.g., naming to definition) which can lead to potential confounds in people with aphasia (for an attempt to circumvent this issue, see for example, Franklin et al., 1995). Nevertheless, participants with aphasia have been found to respond more slowly and/or less accurately when naming target pictures of lower imageability and/or concreteness (e.g., Nickels & Howard, 1995). Imageability also been found to predict the occurrence of semantic errors (Nickels, 1995) although for some individuals this was restricted to targets of low frequency (i.e., a significant interaction between imageability and frequency, Nickels & Howard, 1994).

Effects of imageability/concreteness on word production in aphasia have often been associated with semantic impairments and/or interpreted as being indicative of semantic impairment (e.g., Bird et al., 2000; Howard & Gatehouse, 2006; Luzzatti et al., 2002; Nickels, 1995; Nickels & Howard, 1994). However, they have also been observed in people with aphasia without semantic impairment (e.g., Franklin et al., 1995) and in neurotypical participants (e.g., Barry et al., 1997; Ellis & Morrison, 1998). For example, Franklin et al. (1995) report the case of DRB, a man with aphasia and intact comprehension for written abstract words, but poor production of these items (e.g., reading aloud of irregular abstract words, verbal fluency with abstract words).

Reverse imageability and concreteness effects.

While the vast majority of the literature reports better processing for high imageability and/or concrete words than words low in these features, there have also been reports of the reverse pattern. While much of this research has focused on comprehension (e.g., Breedin et al., 1994; Crutch and Warrington, 2005; Macoir, 2009; Reilly et al., 2007), there have also been reports of this pattern in production (e.g., Gvion & Friedmann, 2013; Marshall et al., 1996; Sirigu et al., 1991; Warrington & Shallice, 1984). For example, Gvion and Friedmann (2013) report the case of Nissim, a man with post-stroke aphasia who showed impaired comprehension and production of imageable concepts but spared ability to comprehend and produce abstract words. In common with most of the individuals reported, Nissim's impairment pattern was attributed to a semantic impairment, in his case with a focus on impairment of visual features (see also Breedin et al., 1994). Similar patterns have been reported by, for example, Marshall et al. (1996) who found that RG, a man with jargon aphasia, was generally poor in comprehension and production of concrete words with abstract words relatively preserved. Interestingly, he showed significantly improved performance in naming concrete words in response to abstract definitions compared to from concrete definitions. For example, naming the target 'medal' in response to either a concrete definition 'a gold disc with a ribbon worn on the chest' or an abstract definition 'an award for valour'. In another example, Crutch and Warrington (2003) attribute the preservation of propositional speech in a man with anomia to relatively intact word retrieval for abstract words relative to concrete words.

Theoretical accounts of imageability and concreteness effects in aphasia.

As will become a recurring theme across the different semantic variables, the mechanism underlying the effect of imageability and/or concreteness on naming performance, remains an area of discussion. It is certainly the case that few authors would claim that the mechanism for imageability actually relates to the ability to conjure up an image. Indeed, Baddeley et al. (1982) argued that although imageability had clear effects (in their case on reading), as a variable it had virtually no explanatory power. This is despite the fact that, as noted above, Paivio (1971) suggested imageability related to how easily a word could be encoded by the nonverbal (imaginal) component of the representational system (in the Dual Coding Theory, see Paivio, 2010, for a review).

The definition for concreteness relating to 'availability to sensory experience' does seem to align better with some accounts of this variable's effects. For example, both Gvion and Friedmann (2013) and Marshall et al. (1996) suggest that the reverse abstractness effect observed was due to specific impairments in visual semantic features, which disproportionately affected processing of concrete words. Kousta et al. (2011) proposed an alternative account for a relative advantage for abstract words when matched to concrete words for imageability in a lexical decision task with neurotypical participants. This was based on the fact that, statistically, abstract words are more emotionally valenced than concrete words, with their hypothesis being that stronger emotionality assists "learning". This is an important area for future research in aphasia (see below for further discussion regarding processing of emotion words).

The fact that concrete words have greater associations with sensorimotor experience has led to the idea that these words have richer semantic representations (e.g., Paivio, 1986). For example, in their computational model, Plaut and Shallice (1993) implemented abstract words as having fewer features than concrete words. Semantic richness has, in turn, been argued to explain why concrete/imageable words are more resistant to impairment (e.g., Nickels & Howard, 1995; Plaut & Shallice, 1991, 1993). However, Franklin et al. (1995) urged caution, citing Paivio et al. (1968) observation that abstract nouns have richer associations, and more different meanings (with the relevant meaning being context dependent; Schwanenflugel & Shoben, 1983). A higher number of associations for abstract words was also reported by Altarriba et al. (1999) who found that participants were able to generate significantly more associations for emotion and abstract words compared with concrete words. Also of relevance here, is a position put forward by Crutch and Warrington (2005; see also, Crutch and Warrington, 2010; Crutch & Jackson, 2011; Crutch et al, 2009; Warrington, 1981) that concrete and abstract words may have different representational frameworks, with concrete words being organised by category, while abstract words are dependent on associative information (that is, words are linked by association while not sharing a category nor being organised hierarchically). Thus, the differences between abstract and concrete nouns may lie in differences in the nature, rather than amount, of the information used in their encoding and therefore may be better captured qualitatively, rather than quantitatively in terms of differences in numbers of features. That is not to say that semantic richness, operationalised as number of features, may not be an important aspect of meaning (see section below) but rather that it does not suffice to account for effects of concreteness/imageability (for discussion in the context of reading, see Recchia & Jones, 2012).

Finally, as noted above, there has been a tendency to conflate imageability and concreteness or to use just one of these measures in the literature. We are aware that we have continued this tradition here. This is a sign of, in part, the statistical difficulty in disentangling effects of these variables given the risk of multicollinearity in analyses. However, it also reflects the ongoing theoretical debate as to the extent to whether imageability and concreteness tap different constructs, and precisely what these constructs might be (see discussions in Connell & Lynott, 2012; Kousta et al., 2011; Paivio, 2013; Warriner et al., 2013). One problem is that as rated variables one can never be entirely sure exactly what individuals are rating. Hence, whilst some authors may believe that these are conceptually different, there is no guarantee that participants are necessarily rating different experiences, and these ratings may indeed be 'contaminated' by other factors including familiarity with the concept.

We would suggest that rather than to attempt to dissociate effects of rated imageability and concreteness, a perhaps more fruitful way forward would be to consider a cohort of alternative measures that together capture the critical constructs of each factor. Ideally these would be objective measures, but alternatively could be ratings of subcomponents, such as association with visual, auditory or motor properties (e.g., sensory experience ratings, Juhasz & Yap, 2013; body-object interaction ratings, Pexman et al., 2019; ease of predication, e.g., Jones, 1985, but see de Mornay Davies & Funnell, 2000; context availability, e.g., Kousta et al., 2011; arousal, e.g., Kousta et al., 2011; Warriner et al., 2013). Indeed, this has been the approach that our team has been taking more recently—rather than focusing on a single 'broad based' factor that might influence semantic processing, we have been examining effects of a range of semantic variables. Each of these variables could plausibly influence spoken word retrieval in picture naming and we hoped to determine which of these potential factors are the most potent for both neurotypical speakers and those with aphasia (e.g., Hameau, Biedermann, et al., 2019; Hameau, Nickels, et al., 2019; Lampe, Hameau, & Nickels, 2022b; Lampe et al., 2021).

We now discuss some of these factors.

Number of semantic features

As noted above, number of semantic features has been taken as an indication of the richness of the semantic representation of the target word. This measure is almost always retrieved from feature norm databases (e.g., Devereux et al., 2014; McRae et al., 2005) where participants are asked to generate features of the concept that a word refers to. For example, in the McRae et al. database, the concept 'cat' is represented by the features *has fur, an animal, a pet, eats, has whiskers, meows, purrs, has four legs, has legs, has a tail, has claws, is domestic, a baby is a kitten, a feline, eats mice, has paws, is independent, a mammal, kills*, and *has eyes*.

For neurotypical participants, higher numbers of semantic features have been found to result in greater accuracy and faster latencies (Lampe, Hameau, & Nickels, 2022b; Rabovsky et al., 2016, 2021; Taylor et al., 2012) in standard, simple picture naming. However, Lampe, Hameau, and Nickels (2022a) found no effect on accuracy in speeded naming for participants who had already completed one round of standard naming on the same items (reported in Lampe, Hameau, & Nickels, 2022b).

Given the prevalence of semantic richness (in terms of number of features) as an explanation for effects of imageability/concreteness on processing, it is surprising that there is only one paper that examines effects of number of semantic features on object picture naming in people with aphasia (Lampe et al., 2021; but see Breedin et al., 1998, for verbs). When number of semantic features was the sole variable in the analysis, Lampe et al. found a significant facilitatory effect on naming accuracy for a group of 175 participants with aphasia, replicating the facilitatory effects reported for neurotypical participants (Lampe, Hameau, & Nickels, 2022b; Rabovsky et al., 2016, 2021; Taylor et al., 2012). However, when other (semantic and lexical) variables were controlled in the analysis, number of semantic features was no longer significant.

Lampe et al. (2021) note that this group of participants with aphasia consisted of individuals with heterogeneous impairments and suggested that this led to considerable inter-individual variability in the strength (and, to a lesser extent, the direction) of the effect of number of semantic features. This was supported by a significant by-participant random slope of number of semantic features: The effect of number of semantic features was (numerically) facilitatory for most participants, but not for all. Given that semantic variables are hypothesised to have their effects at a semantic or lexical level (rather than post-lexical), Lampe et al. carried out additional analyses on a sub-group of 60 participants with primary impairments at these levels (identified on the basis that they made few phonological errors). In this subgroup there was a significant facilitatory effect of number of features (even when other semantic and lexical variables were controlled) and no random slope was retained in the best fitting model, suggesting that the effect of number of semantic features (and the other variables) was similar across participants. Lampe et al. also examined effects of number of semantic features on error patterns, and found that, for these participants with primarily lexical and/or semantic impairments, semantic errors (rather than correct responses) were more likely on words with fewer semantic features.

In terms of mechanisms for these effects, Lampe et al. (2021) argue that the more semantic features that are active, the greater semantic activation overall (e.g., Rabovsky & McRae, 2014, simulation 2). This increased semantic activation is hypothesised to result in increased activation of the target's lexical representation. Therefore, in an impaired language system, this additional activation makes it more likely that the target will be sufficiently active for selection. However, for words with fewer features, the sparse semantic information available may not always result in sufficient activation for a semantic or lexical representation to be selected among similar alternative representations, leading to semantic errors.

In sum, it seems that number of features has a facilitatory effect on word production both in aphasia and neurotypical speakers even when other semantic variables are controlled (Lampe, Hameau, & Nickels, 2022b; Lampe et al., 2021; Rabovsky et al., 2016, 2021; Taylor et al., 2012). However, other research has focused on other aspects of featural representation such as numbers of particular types of features (e.g., perceptual or functional features; e.g., in ageing and Alzheimer's disease, Rico Duarte & Robert, 2014; proportion of features that are visual features, Clarke et al., 2013; number of distinguishing features, e.g., Miozzo et al., 2015), this would seem an important avenue for future research.

Typicality

Typicality (or prototypicality) indexes an item's representativeness within its semantic category. For example, 'robin' is a typical bird, while 'ostrich' is an atypical bird: robins have many features that are prototypical of the category 'bird' (*it is small, flies, has wings and a beak*, etc., e.g., Rossiter & Best, 2013) while ostriches have comparatively few of these features. To operationalise typicality, researchers have most often used ratings of a target's typicality in its semantic category (e.g., Rossiter & Best, 2013) or other measures that have been argued to capture typicality (e.g., frequency of instantiation, e.g., Barsalou, 1985; category potency, e.g., Battig & Montague, 1969; dominance of the category superordinate, e.g., Ashcraft, 1978). However, there are also more objective approaches to capture typicality based on semantic feature norms (e.g., family resemblance score, Rosch & Mervis, 1975).

In neurotypical participants, higher typicality is generally considered to facilitate picture naming latency and/or accuracy (e.g., Dell'Acqua et al., 2000; Holmes & Ellis, 2006; but see e.g., Lampe, Hameau, & Nickels, 2022b, for null effects).

In participants with aphasia, typicality has been most commonly examined using categorisation or semantic decision tasks (e.g., Grober et al., 1980; Kiran et al., 2007; Kiran & Thompson, 2003; Obermeyer et al., 2021; Riley & Thompson, 2010; Sandberg et al., 2012; Stanczak et al., 2006), where it facilitates performance. However, typicality effects in aphasia have also been found in tasks that require verbal output, such as category-exemplar generation (e.g., Grossman, 1981; Hough & Pierce, 1989). There are four relevant picture naming studies in aphasia (Laiacona et al., 2001; Lampe et al., 2021; Rogers et al., 2015; Rossiter & Best, 2013) as well as studies investigating Semantic Dementia (now usually known as semantic variant Primary Progressive aphasia; Rogers et al., 2004, 2015; Woollams, 2012; Woollams et al., 2008), and Alzheimer's disease (Grossman et al., 1998; Morelli et al., 2011).

In the studies with Semantic Dementia, higher typicality items have been reported to result in higher naming accuracy. However, more semantic errors are generally reported to occur on high typicality items, with more omissions on low typicality items (Rogers et al., 2015; Woollams, 2012; Woollams et al., 2008). In stroke aphasia, the results have been more variable. Rossiter and Best (2013) found that higher typicality resulted in higher naming accuracy for a group of 20 people with aphasia (both using matched sets and analysed using regression). Fifteen of their 20 participants showed numerically better performance for higher typicality items (in matched sets) but only two individual participants showed significant effects (with five individual participants showing an effect of typicality in the regression analysis). Moreover, Rossiter and Best noted no clear correspondence between the level of the participant's impairment and the direction and/or strength of the effects of typicality. Laiacona et al. (2001) also only found effects of typicality on naming accuracy for a small number (5) of their 49 participants with aphasia but did not report the direction of the effect. Rogers et al. (2015) found no evidence for significant effects of typicality on naming accuracy for a group of people with aphasia and semantic impairments, with visual inspection of the participants' performance suggesting individual differences with facilitatory, inhibitory, or null-effects across participants.

Lampe et al. (2021) also found no significant effects of typicality on naming accuracy for either their full group of people with aphasia or a subgroup with predominantly semantic and/or lexical impairments, nor was there any evidence of significant variability (i.e., random slopes for participants were not retained for typicality). Lampe et al. noted that these findings held regardless of whether typicality was entered as a sole variable in the analysis or in combination with other potentially influential psycholinguistic variables (including semantic variables). Indeed, a Bayesian correlation found substantial evidence *against* an effect of typicality on accuracy being present (BF_{01} = 6.00). Lampe et al. did, however, find an effect of typicality on the likelihood of semantic and/or coordinate error over a correct response in the full participant group and the subgroup, with more errors on words with higher typicality. Some of these effects interacted with the degree of semantic impairment and were largest in participants with unimpaired semantics. These findings were interpreted as suggesting that for words with higher typicality their standard configuration of semantic features leads to co-activation of incorrect candidate names, which, when selected, result in semantic (or coordinate) errors. Alternatively, within a different processing architecture, when attempting to settle in the attractor of a highly typical item, the impaired system may be diverted to an attractor of another typical representation, which is closely clustered in semantic space, resulting in a semantic error (e.g., Rosch, 1973; Rosch & Mervis, 1975; Woollams et al., 2008).

In sum, the effects of typicality seem less consistent than one might predict, and there are complex interactions with level and degree of impairment. It is also of note that descriptions of the potential mechanisms underpinning effects of typicality differ. For example, Woollams (2012; Woollams et al., 2008) argues that typicality effects arise from the frequent co-occurrence of features in highly typical items. For example, typical birds will tend to have *be light, have wings*, and *be able to fly* with these three features often co-occurring and in some computational models (e.g., Rogers et al., 2004), it is this co-occurrence that drives the superiority of more typical words.[8] In contrast, other theories suggest that rather than the similarities, it is the relative distinctiveness of concepts that is important. For example, in Plaut (1996) connectionist model of semantics, lesioning resulted in more impaired performance for typical than atypical words. Plaut suggests this pattern is due to the network finding it easier to distinguish between atypical words as they have fewer close neighbours due to their distinguishing features. Together these accounts allude to three other potential influences on performance, all of which can be measured independently: intercorrelational density, distinctiveness and number of close semantic neighbours. We will discuss these in turn.

Intercorrelational density

McRae et al. (1997) suggested intercorrelational density[9] as a means of capturing the relationship

[8] Note that Rogers et al. (2015) argue that in Semantic Dementia the cooccurrence of features provides a typicality advantage, due to the hypothesised loss of distinguishing features. In contrast, in 'semantic aphasia' typical items create more competition leading to poorer performance as greater 'semantic control' is required to distinguish between these items and 'semantic control' is hypothesised to be impaired in this population.

[9] A related yet different measure is *intercorrelational strength* (e.g., McRae et al., 1997, 1999), which captures the degree to which a specific feature (e.g., *has fur*) is correlated with the other features of a concept (e.g., 'cat'). Intercorrelational

between all the semantic features of a concept: Some semantic features tend to occur together across concepts, for example, the features *has fur* and *has four legs* occur together across concepts like 'cat', 'dog', 'wolf', 'caribou' and 'cougar' and can be said to intercorrelate.

Stronger correlations between features in words with higher correlational density are argued to result in greater mutual co-activation, due to, for example, bidirectional feature-feature connections (e.g., Cree et al., 1999; McRae et al., 1997, 1999), which allow features in a cluster of intercorrelated features to boost each other's activity. Differences in feature correlations in different semantic categories have also been suggested as one of the reasons for domain-specific impairments in participants with language disorders (e.g., Devlin et al., 1998; Gonnerman et al., 1997).

Intercorrelational density has been investigated in word production studies with neurotypical participants (Lampe, Hameau, & Nickels, 2022b; Rabovsky et al., 2016, 2021; Taylor et al., 2012). Rabovsky et al. (2016) found longer response latency and more naming errors for words with higher intercorrelational density. However, when Rabovsky et al. (2021) asked participants to name the same set of pictures twice, they only found an effect of intercorrelational density on response latencies on the second naming round, although an inhibitory effect on naming accuracy was present in both presentation rounds. Lampe, Hameau, and Nickels (2022a) reported no effect on latency in the combined data from standard and speeded picture naming tasks, however, there was a significant effect on overall naming accuracy, with no evidence for a difference in the effect between the two naming tasks. In addition, Taylor et al. (2012) and Lampe, Hameau, and Nickels (2022b) did not find significant effects of intercorrelational density on standard picture naming accuracy and/or latency. Where effects of this variable were observed, they were interpreted as reflecting enhanced co-activation of semantically related lexical representations, which compete for selection with the target word and slow selection down and/or increase the likelihood of erroneously selecting an incorrect but semantically related representation.

Only Lampe et al. (2021) have investigated effects of intercorrelational density in people with aphasia, and found no significant effects on accuracy (and substantial evidence against an effect BF_{01} = 5.85) or error types in picture naming. Consequently, it seems that this variable has limited effects on word production, despite its plausibility as a factor.

Distinctiveness

In contrast to typicality and intercorrelational density which focus on the similarity between concepts and feature clusters, distinctiveness relates to the importance of differences between features of concepts. There are two different measures: one that focuses on individual features and the other that focuses on concepts.

---

strength determines the level of activation of that particular feature but also of the other features of the concept (e.g., *has four legs, has whiskers, has a tail,* etc.) as features in a cluster of intercorrelated features boost each other's activity. This can affect processing, with more strongly correlated features speeding up activation and thus decreasing processing times. For example, strength of the correlation between a feature and a concept has been shown to predict response times in feature verification tasks where participants were asked to verify features as true or false of a concept (McRae et al., 1997, 1999). For instance, *is hunted* is more strongly correlated with the other features of 'deer' than of 'duck' and was therefore verified more quickly than for the concept 'deer' than the concept 'duck' (for similar findings see also e.g., Garrard et al., 2005; Randall et al., 2004; Taylor et al., 2004).

Featural distinctiveness.

This measure relates to how distinctive a particular feature is. In other words, the degree to which a feature is shared with many other concepts (e.g., *has fur*) or is more uniquely associated with a single or only few concepts (up to 3 in McRae et al., 2005) (e.g., *moos*).

Semantic features that are relatively unique to a concept (distinctive features, sometimes also called distinguishing features) are highly informative regarding that concept and have been found to influence processing in various paradigms (e.g., feature verification, Cree et al., 2006; Taylor et al., 2004), suggesting a privileged role in the computation of word meaning. However, Grondin et al. (2009) showed that lexical and concreteness decisions in neurotypical participants were more strongly facilitated for words with many shared rather than many distinctive features. Vieth et al. (2014) investigated the role of distinctive and shared features in neurotypical participants in the context of the picture-word interference paradigm and found that close distractor-target pairings which did or did not involve distinctive features (e.g., 'horse'—'zebra', with *has stripes* being a highly distinctive feature), elicited the same degree of interference on picture naming. Moreover, when using distinctive and shared features as part-whole distractors, only non-distinctive part-whole distractors slowed naming latencies, which was interpreted as resulting from many co-activated lexical competitors, while distinctive part-whole distractors did not differ from unrelated distractors. However, to our knowledge the only investigation of an effect of number of distinctive features on simple picture naming in neurotypical participants is by Miozzo et al. (2015). They used Principal Component Analysis to combine number of distinctive features and number of encyclopaedic features and found no significant effect of this measure on naming latency (although this measure predicted MEG data as early as 150ms post picture onset in the posterior inferior temporal cortex).

No study has directly investigated the effect of the number of distinctive features that an item has on word production in people with aphasia. However, in a series of studies Mason-Baughman and Wallace (Mason-Baughman, 2009, 2010; Mason-Baughman & Wallace, 2013a, 2013b; Wallace & Mason-Baughman, 2012) investigated the influence of knowledge of distinctive features on word retrieval and word comprehension abilities of individuals with aphasia. In many of these studies, feature distinctiveness was investigated alongside feature importance, which indexes the degree to which features are frequently identified for a concept.

For effects on word production, Mason-Baughman (2009) reported that knowledge of distinctive features, as identified in a feature-word sorting task, did not predict performance on a standardised naming test (Boston Naming Test-2). However, when feature importance was also accounted for, there was a significant correlation between low-importance distinctive feature knowledge and naming performance (see also Mason-Baughman, 2010). Similarly, Wallace and Mason-Baughman (2012) found that knowledge of low- and mid-importance distinctive features was correlated with participants' success in naming pictures of items that contained the features from the feature-sorting task.

It has been suggested that distinctive features may be particularly susceptible to loss during the progression of degenerative conditions like Alzheimer's disease and Semantic Dementia, which has also been proposed to underpin category specific impairments (e.g., Catricalà et al., 2015; Garrard et al., 2005; Gonnerman et al., 1997; Laisney et al., 2011; Moss et al., 1998; Rico Duarte et al., 2009; Rogers et al., 2004; Tyler et al., 2000; Tyler & Moss, 2001; see also Caramazza & Shelton, 1998, for a participant with stroke aphasia). However, investigating naming in individuals with Alzheimer's disease, Garrard et al. (2005) found no clear evidence that knowledge about distinctive information facilitated naming accuracy. Moreover, Sartori and Lombardi (2004) suggested that category specific impairments in Alzheimer's disease can disappear and even be reversed if a related measure capturing the importance of a semantic feature for the meaning of a concept, the relevance of the semantic features, is controlled.

Concept distinctiveness.

This concept-based measure of distinctiveness captures information regarding the whole concept and not single semantic features. In contrast to feature distinctiveness, concept distinctiveness measures how special or informative the features of an item are on average, with higher distinctiveness indicating a higher proportion of distinctive semantic features. For example, 'book' is a highly distinctive concept (based on the McRae et al. (2005) feature norms) with 13 distinctive features including, *stores knowledge, is found on shelves*, or *used by reading*, resulting in a high mean distinctiveness value. In contrast, 'scarf' shares all its features with multiple other concepts in the database and therefore is low in mean distinctiveness.

In neurotypical participants, Taylor et al. (2012) found that higher concept distinctiveness led to faster responses in picture naming. The same pattern was found by Rabovsky et al. (2016), but only when the measure intercorrelational density was excluded from the analysis. In contrast, Lampe, Hameau, and Nickels (2022b) reported the opposite, higher concept distinctiveness slowed response latencies. Similarly, Lampe, Hameau, and Nickels, (2022a) found that distinctiveness significantly slowed responses in both standard and speeded naming with a stronger effect in speeded than standard naming, and no effect on naming accuracy. Lampe, Hameau, and Nickels (2022a, 2022b) acknowledged the difficulty in theoretically explaining these findings and suggested they may stem from enhanced lexical competition from a small number of closely related concepts (e.g., 'kangaroo' when naming the target 'wallaby' that otherwise has rather unique features and therefore high concept distinctiveness). Moreover, in a word learning study with neurotypical participants, Breining et al. (2019) reported contrasting effects of semantic and phonological distinctiveness, arising at different levels of the production system. While semantic similarity during word selection facilitated learning of distinctive features, segmental similarity during segmental encoding facilitated learning of shared segments. Given these findings, it is possible that the lack of consistency in effects of distinctiveness in the word production literature may at least partly be driven by differences in, and lack of control for, phonological similarity of the examined items. In aphasia, once again only Lampe et al. (2021) have directly investigated word production[10] (but see Caramazza & Shelton, 1998, for an investigation in a feature verification task) with no significant effects of distinctiveness found on picture naming accuracy (and substantial evidence against an effect BF_{01} = 5.91), nor the likelihood of a semantic or omission error. Consequently, once again, a semantic variable that has been suggested to be influential in semantic and lexical processing does not have evidence to support a major role in aphasia (and inconclusive evidence in neurotypical participants).

Animacy

There is a long history of literature examining effects of animacy (or the living/non-living dichotomy) on

[10] Humphreys et al. (1988) and Riddoch & Humphreys (1987) did investigate a related concept—the extent to which a category was rated as having shared features between category members (structural similarity of items within a category). Looking at the same single case, both papers reported greater naming accuracy for items from categories rated as having fewer shared features (i.e., higher distinctiveness). However, effects of distinctiveness were not assessed at the item level (only the category level).

performance in aphasia, as one of the most prevalent category-specific patterns of performance (e.g., Hart et al., 1985; Sartori & Job, 1988; Warrington & Shallice, 1984), and effects of animacy have been found in picture naming of neurotypical speeded participants (slower and less accurate responses to living things, e.g., Hodgson & Lambon Ralph, 2008).

However, methodological issues have been prominent in this literature, including lack of control for important variables such as frequency, concept familiarity and visual complexity (e.g., Funnell & Sheridan, 1992; Howard et al., 1995) and problems with control data (e.g., controls at ceiling, Best et al., 2006; Laws, 2005; sex-differences in category knowledge, Laiacona et al., 2006). In 2003, Capitani et al. argued that, of the 79 case studies of category-specific disorders at the time, less than half provided statistically and theoretically informative data. Similarly, Chen and Rogers (2014) note that "the research has followed a characteristic pattern in which potential confounds are identified and subsequently controlled in an effort to reveal the 'true' underlying impairment (...). The confounds continue to multiply, however, so that there may be no 'real' underlying cause beyond the factors reviewed" (Chen & Rogers, 2014, p. 341).

Nevertheless, there is now reasonable consensus that there are cases where performance dissociates between living and non-living things (see e.g., Mahon & Caramazza, 2009, for a review and examples of cases by, e.g., Barbarotto et al., 1995: living things, people, architectural knowledge vs tools, vehicles, furniture; Hillis & Caramazza, 1991: animals vs non-animals; Laiacona & Capitani, 2001: nonliving vs living). Yet, Mahon and Caramazza (2009) argued that the theoretical questions remain the same as those originally debated in the 1980s: "Are different types of information involved in processing different semantic categories and, if so, what distinguishes those different types of information?" (Mahon & Caramazza, 2009, p. 44).

Given the primary focus of this paper is on lexical retrieval and not on the nature of semantic representation, we will not dwell further on these debates here. We note, however, that all current theories regarding animacy effects, assume that concepts are composed of distinct types of information (Mahon & Caramazza, 2009), and many of the accounts attribute different performance to differences in semantic structure across animate and inanimate objects, particularly in terms of intercorrelations among their features, and their distinctiveness (e.g., Gonnerman et al., 1997; McRae et al., 1997; for review see Chen & Rogers, 2014). However, as is clear from our discussion earlier, in general, there is little evidence that these factors (distinctiveness, intercorrelational density) influence picture naming in people with aphasia. This suggests that perhaps the explanatory power of these proposed underlying factors is overstated and the source of animacy effects lies elsewhere.

Semantic neighbourhood

The final semantic variables that we will discuss in detail relate to semantic neighbourhood, the influence of which is probably one of the most prevalent ideas in the discussion of lexical access. Semantic neighbours of a target word are other words that share part of their meaning with the target. The influence of this variable is based on the idea that a cohort of semantic neighbours is activated at the lexical level, and competition between these neighbours influences ease of lexical retrieval. As noted above, it is this concept that underpins much of the research on context effects, where a neighbour (usually an item from the same category) is pre-activated (see e.g., Spalek et al., 2013, for review; Bürki et al., 2020, for a meta-analysis). However, effects of a word's semantic neighbours have also been investigated in picture naming without external manipulations of context.

Semantic neighbourhood density.

Semantic neighbourhood density refers to the number of semantic neighbours of an item, with most authors examining effects of number of *near* semantic neighbours (often from within the same category). While, in neurotypical participants, number of semantic neighbours has been reported to affect processing in a variety of tasks, such as visual lexical decision (e.g., Buchanan et al., 2001; Pexman et al., 2008), no consistent effects have been found on word production: In a speeded picture naming task, Mirman (2011) observed that errors (especially semantic errors) in neurotypical older adults increased for words with many near semantic neighbours but decreased for words with many distant semantic neighbours. Similarly, Fieder et al. (2019) found inhibitory effects of near semantic neighbours on speeded naming latencies and accuracy, as well as an increase in semantic errors and omissions (but no effect of distant neighbours), in neurotypical young adults. Lampe, Hameau, and Nickels (2022a) found that in combined data from standard and speeded picture naming, participants were slower to name pictures with higher numbers of near semantic neighbours.

In contrast to these significant effects of semantic neighbours on speeded naming, standard picture naming studies with neurotypical adults did not yield any effect of semantic neighbourhood density (Bormann, 2011; Hameau, Nickels, et al., 2019; Lampe et al., 2017; Lampe, Hameau, & Nickels, 2022b). However, Lampe, Hameau, and Nickels (2022a) also found no significant difference when comparing the magnitude of the effect between standard and speeded picture naming.

Different directions of effects of near and distant semantic neighbours in neurotypical speakers (Mirman, 2011) have been explained with attractor dynamics, such that near neighbours create competing attractor basins that generate inhibition, while the occurrence of many distant neighbours results in a "pull" towards the target, thereby facilitating naming. Fieder et al. (2019) explained inhibitory effects of near neighbours within the Swinging Lexical Network Hypothesis (Abdel Rahman & Melinger, 2009, 2019) in which naming is characterised by the activation of a cohort of semantically related items at the lexical level: The larger the cohort (i.e., the more semantic neighbours), the more competition, hence the more inhibition of naming performance. Conversely, the absence of semantic neighbourhood density effects in standard picture naming was discussed by Bormann (2011) as consistent with non-competitive accounts of spoken word production (e.g., Dell, 1986), in which the time to reach the threshold which triggers phonological encoding does not depend on the activation states of other words.

One of the complications of this literature, however, is that a variety of different ways of capturing semantic neighbourhood density have been implemented: feature-based (e.g., Mirman, 2011; Mirman & Graziano, 2013), association-based (e.g., Hameau, Nickels, et al., 2019; Mirman & Magnuson, 2006), and context-based (Kittredge et al., 2007b) neighbours, as well as ratings of the number of category coordinates (e.g., Bormann, 2011). As noted above, and consistent with Bormann's ratings, most authors, when considering semantic neighbourhood density, refer to number of *near* semantic neighbours, or contrast near and distant semantic neighbours (e.g., Fieder et al., 2016, 2019; Mirman, 2011). Hameau, Nickels, et al. (2019) attempted to determine whether the use of different measures of neighbourhood density underpinned the variability in the effects found. In doing so, they used principal component analysis and found that six different measures of neighbourhood density loaded onto four different components. They labelled these feature-based neighbours (with strong contributions from near neighbours defined as words with a cosine feature vector similarity of > .4, Mirman, 2011, and rated number of category coordinates, Bormann, 2011), contextual neighbours (contributions from raw contextual neighbours, within category contextual neighbours, Kittredge et al., 2007b), association neighbours (Mirman & Magnuson, 2006) and distant neighbours (distant neighbours defined as words with a cosine feature vector similarity of > 0 and ≤ .25; Mirman, 2011). However, none of these components tapping neighbourhood density significantly predicted picture naming in neurotypical speakers, and they conclude that effects of these factors may be too small to be reliably detected in neurotypical speakers without the manipulation of context.

Effects of near semantic neighbours on naming accuracy in people with aphasia have also differed across studies, with either overall null (Blanken et al., 2002; Bormann, 2011; Bormann et al., 2008; Kittredge et al., 2007b; Lampe et al., 2017; Lampe et al., 2021) inhibitory (Mirman, 2011) or facilitatory effects (Kittredge et al., 2007a). Moreover, both facilitation and inhibition have been observed for measures representing the number of close semantic neighbours across different individuals within a single study (Mirman & Graziano, 2013). In addition, when analysing errors, a high number of semantic neighbours has been shown to increase the chance for a semantic error as compared to a correct response (Mirman, 2011; but see Fieder et al., 2019, and Lampe et al., 2017, for null effects), or to an omission (Kittredge et al., 2007b). Finally, the probability of making an omission relative to a correct response has been found either not to be affected by semantic neighbourhood density (Kittredge et al., 2007a) or to be increased (Kittredge et al., 2007b).

Once again, in order to address the potential source of this variability, Hameau, Nickels, et al. (2019) used the same principal components described above (and those representing lexical and post-lexical factors), to examine the effect of different aspects of neighbourhood density in aphasia. Using data from 193 individuals with aphasia from the MAPP Database (Mirman, Strauss, et al., 2010), no measure had a significant main effect on accuracy. However, there was significant individual variability (significant random slope) for the semantic feature-based neighbourhood component, with supplementary analyses showing that more participants (around 68%) showed a facilitatory effect or trend. Hameau, Nickels, et al. noted that while the different measures used across the previous literature may indeed have influenced the effects found, the participants studied may have been an equally potent factor. Kittredge et al.'s two studies (Kittredge et al, 2007a, 2007b) provide a good example: Kittredge et al. (2007a) reported higher accuracy for words with many contextual neighbours for a group of 50 people with aphasia. But when, in Kittredge et al. (2007b), a further 50 participants were included, this effect was no longer present.

Hameau, Nickels, et al. (2019) were unable to find any participant-related factors that might determine the cause of the individual variability in their data: The individuals who showed the strongest inhibitory effects did not differ from those showing the strongest facilitatory effects on any measure examined (conceptual processing[11], proportion of semantic/mixed errors compared to phonological errors, repetition ability, phonological or semantic weights, Foygel & Dell, 2000, anomia severity, aphasia severity, aphasic syndrome, age, education). However, they suggest that the fact that an increase in the number of feature-based neighbours both inhibited and facilitated performance across different people with aphasia challenged the idea that there is substantial competition between semantically related words in picture naming, when their activation is a natural consequence of lexical activation (as opposed to a result of priming or contextual manipulation). They suggested that the observed effects were most likely to be due to a complex balance between facilitation and inhibition. Facilitatory effects could indeed arise within the conceptual level, from spreading of activation to related concepts, or they could originate from feedback from the lexical to conceptual levels. Inhibitory effects, on the other hand, most likely result from lexical competition. Whether the 'net' effect is facilitatory or inhibitory will be

[11] There was also no interaction between conceptual processing (Pyramids and Palm Trees score) and effect of feature-based neighbourhood in the main analysis.

a function of the characteristics of a word and its neighbours, and of the nature of the impairment (if any).

As noted above, Lampe et al. (2021) also examined the effects of number of near semantic neighbours (from feature overlap), in the context of controlling not only for psycholinguistic variables but also other semantic variables (imageability, number of semantic features, typicality, distinctiveness, intercorrelational density, semantic similarity). Importantly, their dataset overlapped with that of Hameau, Nickels, et al. (2019). Like Hameau, Nickels, et al., they found no significant effects of semantic neighbourhood density on naming accuracy or error type. Indeed, Bayesian correlation showed evidence *against* an effect of this variable (BF_{01}=7.48). However, in contrast to Hameau, Nickels, et al., they also did not find significant variability (the random slope did not improve model fit). Lampe et al. discuss this potentially surprising result, but for our purposes this lack of replication, even with only slight differences in participants, stimuli and measures, serves to underline the fragility of these effects.

Semantic similarity.

Another measure that relates to semantic neighbours concerns the semantic similarity of a word to other words in the mental lexicon ranging from dissimilar with no shared features (e.g., 'basket' and 'ambulance') to very similar with near identical features (e.g., 'blackbird' and 'raven'). In word production, semantic similarity has mostly been examined in context manipulation tasks that contrast the influence of semantically more similar and semantically less similar distractors on target word processing. Distractor words that are semantically more similar to the target have mostly been found to interfere more strongly with production of the target word than distractors that are semantically distant (e.g., Rose et al., 2019; Vieth et al., 2014; Vigliocco et al., 2004; but see Mahon et al., 2007, who contrasted within-category near and far distractor words, for contradictory findings). Moreover, Rose and Abdel Rahman (2017) found that semantic similarity influenced the cumulative semantic interference effect in continuous naming.

There have been very few investigations of this variable without context manipulation in neurotypical participants: Fieder et al. (2019) found a significant inhibitory effect of semantic similarity (average similarity of all of a target's semantic neighbours) on naming accuracy but not latency in a speeded naming task, which was interpreted as indicating enhanced lexical competition between co-activated representations for words with higher semantic similarity. However, this was not replicated by Lampe and colleagues (using average similarity of all concepts in the McRae et al., 2005, database) in standard naming (Lampe, Hameau, & Nickels, 2022b). They also did not find an effect of semantic similarity using data combined across standard and speeded picture naming nor (in Lampe, Hameau, & Nickels, 2022a) a difference in the magnitude of the effect between the two tasks.

To our knowledge, only Lampe et al. (2021) have investigated effects of semantic similarity on naming in people with aphasia, and they found no main effects on accuracy or error types. However, there was an interaction with severity of semantic impairment on the likelihood of semantic errors: Participants with more severe semantic impairments were more likely to make a semantic error on words with lower semantic similarity, while individuals with less severe or no semantic impairment showed little effect of semantic similarity on performance. Lampe et al. suggest that when semantics is intact (in neurotypical participants or those with aphasia), or only slightly impaired, the target may be sufficiently activated relative to any co-activated competitors for word production to be successful irrespective of the target's semantic similarity. However, in the case of severe semantic impairment, when processing words with higher semantic similarity, individuals benefit from greater spread of activation to (and from) similar concepts, which facilitates processing of the target concept. In contrast, for words with lower semantic similarity, there may be reduced spread of activation at the semantic level, which, in addition to corrupted or noisy semantic representations, results in insufficient activation of the target concept. This reduced semantic facilitation may then be outweighed by competition from co-activated lexical representations, resulting in the incorrect selection of semantically related words.

Importantly, Lampe et al. (2021) note the difficultly in interpreting effects of this measure given that semantic similarity of an item can be low because there are either no closely related concepts at all, or because it has few closely related concepts (near semantic neighbours) but many distantly related concepts. Hence, while it seems intuitively important to account for how similar an item is to others in semantic space, how to measure this is less straightforward.

Further semantic properties

We have discussed in some detail the research on the effects of several semantic variables on lexical retrieval. There are several others that we have not described so far, usually because there has been less research in the literature focusing on these or the effects are less robust. We will briefly address some of these here.

Operativity.

Items that are operative are those that are discrete and separate from their context, manipulable, firm to the touch and available to several sensory modalities (e.g., ‘vase’, ‘finger’, ‘fire hydrant’), in contrast to those items that are figurative (continuous with their context, not easily manipulated, difficult to grasp, and known primarily by their visual features; e.g., ‘ceiling’, ‘waist’, ‘road’). Gardner (1973) suggested that operative items benefited from being represented in ‘action schemata’ as well as with multiple sense representations. Although, Gardner showed significant effects of operativity on naming in aphasia, he failed to control for many psycholinguistic variables that we would now consider important. Nevertheless, even controlling for these variables, Nickels and Howard (1995) found that eight of 27 people with aphasia showed significant effects of operativity (and this was maintained for seven participants when animacy was added to the analysis, Howard et al., 1995; see also Rossiter and Best, 2013). Howard et al., using a similar analysis and controlling for the same confounding variables, found effects of operativity on picture naming for only four of their 18 participants in a regression analysis, with two of these showing a reverse effect (more operative items were named more poorly). Interestingly, when Howard et al. attempted to tease apart the effect of operativity by separately examining effects of manipulability, availability to multiple senses and separability from the environment, only the latter was not significant at the group level, but, critically the significant effect of manipulability was in the reverse direction to that predicted[12]—items that were less manipulable were better named (and this accounted for the individual effects of operativity).

While interesting, these studies on operativity once again show the complexity of semantic effects on word production and the clear variability across participants. Moreover, it is apparent that the factors underlying the effects of operativity could be still further dissociated (e.g., by different senses such as visual, tactile) and

[12] This is also broadly related to the literature on visual ‘affordances’ (shapes allow for certain types of manipulations and actions independent of specific knowledge of their identity, Cosentino, 2019; Gibson, 1979) which have been suggested to be intact even in cases of severe conceptual impairment in Semantic Dementia (Hodges et al., 2000).

other factors discussed here also controlled (e.g., number of semantic features).

Abstract concepts and emotion.

Earlier we discussed the role of abstractness and imageability in lexical access and evidence from aphasia. However, the vast majority of research on word production in general and in aphasia in particular, has used data from picture naming, for perfectly understandable practical reasons. Consequently, historically, there was relatively little research that has focused on abstract word production. Where there is research, it has largely depended on tasks that use written word stimuli, however, this is problematic in that any effects of abstractness or imageability could arise in mechanisms specific to processing of visual stimuli. Nevertheless, there has recently been more attention on what influences processing of abstract words.

Abstract words have traditionally presented a conundrum for componential representation accounts: Concrete words have identifiable features, but what are the “features” of abstract concepts? An area of growing interest within this domain has been the role that emotion may play in grounding abstract words in sensorimotor information. Emotion in words is typically defined by two features: valence (how much a word is associated with positivity/negativity) and arousal (associations with activity/passivity). Evidence from neurotypical populations has shown an impact of valence/arousal on a number of language tasks, including response times in lexical decision tasks (e.g., Kousta et al., 2009; Kuperman et al., 2014) and word learning (Ferré et al., 2015; Ponari et al., 2020). Abstract words tend to be ‘more emotional’ (i.e., have more extreme ratings of positivity/negativity) than concrete words (e.g., Kousta et al., 2011) and it would appear that emotion may be particularly important in the representation of abstract concepts. For example, Ponari et al. (2018) found an effect of valence on children’s response times on a lexical decision task for abstract but not concrete words, while Ferré et al. (2015) found that valence had a facilitatory effect on the acquisition of abstract but not concrete words.

Research concerning the influence of valence/arousal on lexical processing in people with aphasia is currently emerging. To our knowledge there are only two studies to date of which only one included a production task. Newton et al. (2020) found that people with aphasia had faster response times in a lexical decision task for both positively and negatively valenced words compared with neutral words and that this effect was evident for all levels of aphasia severity. Mason and Nickels (2021) examined people with aphasia’s processing of non-emotional abstract words and words that specifically refer to emotions across three language tasks: free recall, lexical decision and reading. The outcomes from this study were mixed: From a sample of nine participants, three individuals had higher accuracy on the recall task and faster lexical decision response times for the emotion words compared with the non-emotion abstract words. However, one participant showed the reverse effect with an advantage on both tasks for the non-emotion abstract words.

The effect of valence/arousal on lexical processing in people with aphasia is a field for continuing research. It is also noteworthy that there has been a recent movement away from dichotomy between abstract and concrete words in which it is suggested that abstract words should not be approached as one group (Barsalou et al., 2018; Villani et al., 2019). However, how abstract words can be divided into meaningful subsets of words (with regard to lexical processing), remains a point of discussion. Nonetheless, considering how people with aphasia may respond to language tasks with specific subsets of abstract words may be a fruitful area for future research as this field develops.

Summary & Conclusions

No-one would deny the essential role that semantic processing plays in word production—without being able to identify the to-be-expressed concept, whether this is from a thought, an object in the environment or a picture, no word can be selected and produced. Consequently, we suggested that it was surprising that there had been relatively little research examining the effects of the semantic properties of the lexical items themselves on word retrieval in the absence of contextual manipulation. Where the effects of these properties have been examined, they have often been found to be restricted or inconsistent in speakers with typical language skills. It is possible that the relative lack of attention is a consequence of the 'file-drawer' problem with a publication bias against studies showing null results, together with other methodological challenges that we return to below.

Given our impression that findings were potentially more robust than in neurotypical speakers, we focused on the data from people with aphasia, reviewing the literature on the effects of 12 semantic variables on conceptually driven retrieval of single words (mostly in picture naming). Here, we draw these findings together but, given the complexity of the literature, the reader is strongly advised to refer to the sections above for nuance. Lampe et al. (2021; Table 1, and Table A1) also provides a summary of the previous literature on effects of (feature-based) semantic variables in aphasia. Overall, just as in unimpaired speakers, the data from people with aphasia does not result in a clear pattern, and there are likely to be both theoretical and methodological contributors to the effects being limited and inconsistent.

Overview of findings

Of the 12 variables we examined, three variables (valence, arousal, number of distinctive features) have not, to date, been studied on picture naming in people with aphasia. All of the remaining nine variables had been reported to influence picture naming in participants with aphasia in at least some previous studies.

As highlighted above, it was common for the effect of the same variable to be inconsistent across studies or participants with aphasia. For the variables concreteness/imageability[13], typicality, semantic neighbourhood density, and operativity there are reports of facilitatory, inhibitory and null-effects. Similarly, dissociations in performance between living and non-living things were found with individuals with aphasia showing better performance for living or non-living items, respectively.

Four variables (intercorrelational density, semantic similarity, concept distinctiveness, number of semantic features) have only been examined in one study with people with aphasia (Lampe et al., 2021). For intercorrelational density, concept distinctiveness, and semantic similarity there was substantial evidence against a main effect of these variables (Lampe et al., 2021). In contrast, number of semantic features was found to have a facilitatory effect on picture naming of people with and without aphasia. However, even the effects of this variable are relatively fragile.

We noted earlier that semantic variables can have an impact on lexical activation, consequently we would not necessarily expect effects of these variables to be restricted to individuals with semantic impairment.

[13] We report imageabiity and concreteness together here, as they are so highly intercorrelated and few studies have attempted to dissociate the two.

Nevertheless, an interaction with semantic impairment has been noted for some factors, perhaps because the otherwise subtle effects are magnified in these individuals. These variables include imageability, number of semantic features, typicality and semantic similarity.

So where does this leave us? It is undoubtedly the case that factors related to semantic representation have an influence on word retrieval in aphasia, and in neurotypical speakers. However, precisely which factors these are remains uncertain. Unfortunately, we do not feel that we can categorically conclude that variables A & B *do* have an effect on conceptually driven word retrieval, while variables Y & Z *do not*. We would tentatively suggest that number of semantic features is worth investigating further in both people with aphasia and neuroptypical speakers, but even here greater nuance is required (see below for further discussion).

Theoretical Considerations

In theoretical terms, the effects of the semantic variables are generally discussed in terms of the nature of semantic representation and of activation spread within the semantic system, resulting in increased activation of the target at the lexical level. Depending on the theoretical stance taken, there are influences from competition within the semantic system and at the point of lexical selection. However, as we discuss at length elsewhere (e.g., Hameau et al., 2019; Lampe et al., 2021, 2022b) it is extremely complex to predict the effects of the balance of different variables, and consequently the precise theoretical mechanisms underpinning the effects. Some variables may result in increased activation of the target, while others may increase the activation of other similar words in the lexicon. Alternatively, a single variable may increase activation of both target and potential competitors, with the net result (facilitation, inhibition or no net effect) depending on exactly the items in the stimulus set or their other properties (e.g., frequency, or difference in frequency between target and neighbours). For example, as noted above, Nickels and Howard (1994) found that there was an interaction between imageability and frequency, with effects of imageability on semantic errors only being found for items of low frequency. Similarly, it is possible that rather than the number of semantic neighbours, it may be the frequency of those neighbours or the relative frequency of the target and its (closest) neighbours that influences processing (see e.g., Hameau et al., 2021, for a demonstration of this with regard to effects of phonological neighbourhood).

Another possible factor is that variance induced from 'downstream' processes and influences may be more potent and 'mask' the earlier effects of semantic representations on lexical access. For example, if there is a small influence of a semantic factor on the activation of a lexical item, but a large influence of frequency or phonological neighbourhood density, the effect of the semantic factor may not be apparent, or as noted above, only in the interaction of the two factors. But, of course, this does not mean that these factors are not important in terms of the conceptual influence on word retrieval.

Another important issue is that there may be interactions between the influence of types of representation and context. For example, could it be that when naming pictures, visual features will be more prominent compared to when naming from definition? If so, then would items with greater numbers of visual features show a benefit in picture naming that they wouldn't in naming to definition? Similarly, it is possible that associative features (e.g., 'found on a farm') may play more of a role in thematically driven contexts? Oppenheim and Nozari (2021), albeit in the context of blocked cyclic naming, used computational modelling to examine this issue. They assumed that under normal circumstances thematic information is not activated in picture naming. However, in their computational model, when there is a supporting context for naming (i.e., it is blocked thematically; e.g., farm: cow, tractor, barn), an extra thematic feature is activated. This increases the overall lexical activation and facilitates naming relative to when there is no supporting context. This is (yet) another interesting avenue for further research.

We have suggested several potential reasons why the variables that have been examined to date may indeed be theoretically relevant even if we have failed to observe their effects reliably to date. However, it is also possible that an absence of effects may reflect that our conceptual systems do not code those aspects of meaning that we have examined to date. It is certainly the case that theoretical accounts of semantic representation and their relationship with lexical retrieval remain complex. Consequently, predictions are not always clear regarding which aspects of meaning will be important for lexical retrieval, nor indeed which aspects of similarity and differences in meaning are coded in conceptual representations.

As noted above, our ability to predict the effects of variables is likely to benefit from efforts to perform computational modelling of the effects of different aspects of semantic representation on lexical. Nevertheless, this is far from straightforward. For example, computational models using small vocabularies (e.g., Dell et al., 1997; Oppenheim et al., 2010; Oppenheim & Nozari, 2021; Plaut & Shallice, 1993) are incredibly useful for understanding the dynamics of lexical retrieval and its breakdown in aphasia. However, they seem unlikely to be able to adequately allow full exploration of the complex patterns of semantic relationships within and across categories and their effects on lexical retrieval, especially when also needing to take into account the effects of factors at other levels. Nevertheless it is critical that this is attempted, and the use of systematic exploration of the effects of different representational architectures and processing constraints would undoubtedly move the field forward (for an example of this approach see, Rapp and Goldrick, 2000).

Methodological Lessons

We also believe that this review has highlighted some methodological issues that are important for researchers to consider and may have contributed to the inconsistent findings. In this section, we discuss three different issues.

Design and analysis

Earlier, we cautioned that even within 'simple' picture naming tasks there were a variety of methodologies that had the potential to influence results. These include familiarisation with the pictures and their names, naming of the targets more than once (repetition), or speeding of responses. Despite some work examining differences across these task variants, we cannot yet be certain of the extent to which these manipulations affect results and further systematic research comparing effects of semantic variables in different task variations is needed. Consequently, authors need to be cautious when comparing studies or performing meta-analyses that they are indeed comparing like with like. The same is true with reported effects of variables, where the same label (e.g., imageability, semantic neighbourhood density) may be used to refer to the results of different measures (e.g., different rating instructions, or different methods of obtaining objective counts) and these measures cannot be assumed to be indexing the same construct.

In terms of analysis, it is well understood that many factors influence performance on linguistic tasks, and word retrieval in particular (see e.g., Alario et al., 2004; Nickels & Howard, 1995; Perret & Bonin, 2019). It is also well known that misleading results can be obtained if critical variables are omitted from analyses (e.g., Carroll & White, 1973; Howard et al., 1995): One variable may be mistakenly perceived as being important when in fact it is another, with which it correlates that actually influences performance. Alternatively, the effects of a particular variable can be masked by another variable—for example, if an individual shows a strong effect of length, or phonological neighbourhood density on performance, then this may mask effects of semantic variables. Given this, it is of concern that many studies have focused on a single semantic variable, and, in some cases also control inadequately for other non-semantic variables. As already discussed, this has the consequence that it is difficult to be confident that it is indeed this variable that is important and not another (uncontrolled) semantic (or other) factor.

However, controlling for the myriad of variables that are considered important, or potentially important, is no easy task. We have discussed the barriers to using factorial designs given the difficulty of matching for large numbers of variables, and especially when these variables intercorrelate. Because some semantic variables measure similar facets of meaning (e.g., imageability and concreteness), the potential for multicollinearity is particularly high. This both means that is it vital to tease apart the effects of different variables and that this is particularly hard to achieve either in factorial designs or statistically in continuous designs.

Whatever design decisions are made, it is most important that, when using non-factorial designs and regression techniques, at the very least, researchers examine their stimuli for multicollinearity across the widest possible range of variables to determine the extent to which variables intercorrelate. We have suggested the use of regression techniques such as (generalised) mixed effects models or logistic regression. However, within this analysis approach, we have advocated for the use of Variance Inflation Factors, which diagnose when multicollinearity becomes problematic within a model. An alternative is to use Principal Component analysis, if this still allows the research question to be addressed. It would not, for example, be an appropriate technique if one was interested in teasing apart the different variables as we have been here as the resulting components cannot be directly mapped onto the measures from which they are derived. Whatever statistical method is chosen, it is important to be mindful of the implications of that choice and to explicitly mention the limitations of the conclusions that can be drawn, to alert the statistically less skilled reader.

In addition, as noted above, there is the very real possibility that there are interactions between variables. Consequently, in an ideal world, one would not just examine the variables themselves, with suitable control, but also interactions between these variables. This would potentially reveal effects like those for imageability and frequency (Nickels & Howard, 1994), or phonological neighbourhood and frequency (Hameau et al., 2021).

To be able to perform sophisticated statistical analyses using large numbers of variables and also examine their interactions requires a substantial body of data. If not the models may fail to converge or, even if they do, the results may be unreliable. Indeed, the literature has suffered from studies using relatively small numbers of stimuli, which is likely to have contributed to the variability in results. Consequently, there remains a critical need for studies to utilise large numbers of stimuli, particularly when aiming to examine effects at the level of the individual in order that the findings are more robust and all the relevant factors can be examined.

Dissecting meaning

We have already alluded to the fact that many of the variables that have been examined to date (e.g., imageability, concreteness, operativity, animacy, number of semantic features) are rather broad in their scope, hampering interpretation of findings. It is possible that greater understanding will come from trying to 'deconstruct' variables into the different aspects of meaning that underpin these broader concepts. For example, as discussed above, the literature on animacy effects now usually takes this approach. Similarly, while number of features may play a role, it may be the different types of features or the relative balance of different types of feature that is of more importance. In addition, the way that some variables are measured is less than straightforward, and may lead to different potential effects being conflated. For example, semantic similarity has been generally measured using the mean number of shared features between a target and the other words in the lexicon. Yet this conflates the possibility that effects are driven by a single concept that is very similar to the target while most other concepts are highly dissimilar, with, in contrast, effects driven by a large number of moderately similar concepts (the same issue also applies to measures of number of semantic neighbours, for example). This measure also does not take into account whether some types of featural overlap may be more important than others.

There is no easy answer to the difficulties in measuring the facets of meaning, but researchers should do their best to ensure that measures are as reliable as possible by using larger databases, replicating measures (e.g., feature counts, imageability ratings) across databases, and when generating their own ratings, ensuring ratings are obtained using as many participants as possible. Researchers also need to be mindful to be explicit regarding their methods, and that their labelling of factors does not obscure differences. This will assist in the avoidance of further confusion, such as has occurred with, for example, image variability being labelled as, and conflated with, imageability.

Challenges of aphasia

We have focused our review on the patterns found in people with aphasia, given our impression that effects were far from robust in the neurotypical population. However, it is important not to underplay the challenges encountered when using data from people with aphasia to address theoretical questions. Most of the research we have reported used one particular task, picture naming, and aimed to use this to inform our understanding of conceptually driven word retrieval. However, the brain lesions that cause aphasia often result in impairments beyond language, that may introduce additional ‘noise’ into the data.

For example, while most researchers will, in general, have excluded individuals with severe motor speech disorders (dysarthria, apraxia of speech), mild impairments may still be present and impact on overall accuracy and reaction time. One way to attempt to overcome such an impact is to perform analyses examining the effects on different error types (compared to correct responses). However, this approach does not solve the problem for analyses of response latencies. An alternative is to exclude individuals who produce speech sound errors, thereby hoping to restrict the analysis to those individuals with semantic or lexical impairments. This method is suitable for both reaction time and accuracy analyses, but, given the prevalence of such errors in individuals with aphasia, can lead to a catastrophic loss of participants. For example, in Lampe et al. (2021) even using a relatively lax criterion of excluding individuals who produced 5% or more phonological errors, only 34% (*n* = 60) of the 175 original participants were eligible for inclusion in this subgroup.

There are also other, less commonly addressed, potential impairments that may impact performance. For example, Hillis and Caramazza (1995) found that semantic errors in picture naming not only occurred in individuals with lexical or semantic impairments but also in those with an impaired ability to access semantic information from visual stimuli. However, very few of the reported investigations perform sufficient assessment to exclude this impairment as a contributory factor in impaired picture naming. This reiterates the importance of thorough assessment of participants with aphasia, above and beyond the particular task of interest. This is essential in order that those individuals with potential ‘confounding impairments’ can be excluded (for a discussion of this issue in the context of case series research see, Nickels et al., 2011).

There is also the question of which level of impairment is the most relevant for the investigation of the effects of semantic variables on word production. Some might consider that individuals with semantic impairments might be the most appropriate, given that these variables aim to tap aspects of semantic representation. However, individuals with intact semantic processing have also been found to be affected by these variables in ways that provide insights into language processing (e.g., imageability in Franklin et al., 1995). Hence, we have argued that data from individuals with impairments at both semantic and lexical levels could be informative. Once again, it is difficult to be confident whether different levels of impairment might result in different effects of semantic (and other) variables. However, Lampe et al. (2021) found that there were no interactions between effects of semantic variables on accuracy of picture naming and the degree of semantic impairment (as measured by a picture association task), suggesting similar direction of effects on accuracy across participants at least for the variables they studied. Nevertheless, there is the potential for effects of variables at one level of processing to mask effects of variables that arise at another, and for different patterns of errors to occur depending on the level of impairment (e.g., Rapp & Goldrick, 2000). Consequently, it is important that future research ensures that there is also attention to individual patterns (see e.g., Figure 2, Hameau et al., 2019) and/or the effects of different impairments on processing.

Future Directions

In sum, we have reviewed the (complex) literature on the effects of aspects of meaning on lexical retrieval in aphasia. Using data from people with aphasia is particularly helpful given the limited effects of these factors in speakers without language impairment, but is not without challenges itself. Although the picture remains far from clear, we hope that researchers will use the pointers and guidance provided here and take up the challenge to progress the field further. We suggest that it is critical that future research ensures that both non-semantic, and other semantic variables are controlled in the analysis, and have suggested that regression techniques are one way to approach this. Critically, however, the potential for multicollinearity should be at the forefront of researchers' minds with the correlations in the data investigated and the effect on the model monitored using, for example, variance inflation factors. It is also critical that large numbers of stimuli are used in order that the data as reliable as possible. In addition, given the inevitably (and increasingly) large number of factors that need to be included in the model structure, a large number of data points are required in order that the model will converge and there is not 'overfitting'.

We have also suggested that the relative lack of clear effects in the literature to date could result from the effects of variables being interactive, with some evidence already emerging that this may be the case. Consequently, exploring interactions between variables in statistical models is also important as this field moved forward. In addition, it remains unclear which aspects of meaning are important and relevant. There is no easy solution to this problem, except through the systematic investigation of the different, theoretically likely, candidate variables, and building on the knowledge gained from previous research (including that reviewed here). We certainly do not mean to suggest that the variables reviewed here are a 'gold standard' list, nor that one type of variable (e.g., feature-based) is preferable to another.

One could ask whether one semantic variable could be thought to be more clinically relevant than another and this could be a potential way of prioritising and focusing research? While this would seem sensible, in order to determine whether a variable is important for clinical practice, one first needs to discover whether it has a reliable and consistent effect on performance in the population in general—it is those variables that markedly affect performance that will be more clinically relevant. Consequently, the research will drive our understanding of clinical relevance rather than vice versa. For example, if the apparently widespread difficulty with abstract words is discovered to be attributable to one particular facet of their meaning (rather than 'abstractness' itself) then words with this facet should be assessed and, if influential for this individual, targeted in treatment—rather than all abstract words. Along those lines, we have proposed that many of the variables examined to date may remain too broad (e.g., concreteness, animacy, number of features) and it is important to attempt to tease apart which aspects of the many ways that, for example abstract and concrete word differ, are those that underpin performance.

Finally, as semantic representation and processing is so multifaceted, the impact of each facet on processing is hard to predict. Consequently there is an urgent need for computational modelling that systematically investigates how different aspects of meaning representation affect word retrieval under different processing conditions in both neurotypical and impaired language production. For those of us who lack the skills to programme computational models, it would also be incredibly valuable for such models to be accessible to others for them to simulate their data, in a similar way to Webfit (http://langprod.cogsci.illinois.edu/cgi-bin/webfit.cgi) that allows researchers to examine the effects of different lesion types on naming performance within Dell et al.'s (1997) model of word production. Such tools should help us towards the goal of better understanding semantic representation, its impact on word retrieval and its breakdown in aphasia. We would hope that this understanding will in turn enable clinicians to better focus of assessment and treatment.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713–734. https://doi.org/10.1080/01690960802597250

Abdel Rahman, R., & Melinger, A. (2011). The dynamic microstructure of speech production: Semantic interference built on the fly. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(1), 149–161. https://doi.org/10.1037/a0021208

Abdel Rahman, R., & Melinger, A. (2019). Semantic processing during language production: An update of the swinging lexical network. *Language, Cognition and Neuroscience*, *34*(9), 1176–1192. https://doi.org/10.1080/23273798.2019.1599970

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531–552. https://doi.org/10.3758/BF03200732

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155. https://doi.org/10.3758/BF03195559

Allison, P. (2012). *When can you safely ignore multicollinearity?* Statistical Horizons. https://statisticalhorizons.com/multicollinearity

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (Eds.), *Current Perspecrives in Dysphasia* (pp. 32–60). Churchill Livingstone.

Altarriba, J., Bauer, L. M. & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31,* 578–602. https://doi.org/10.3758/BF03200738

Ashcraft, M. H. (1978). Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 155–164. https://doi.org/10.1016/S0022-5371(78)90119-6

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers 1*, 1–45.

Baddeley, A. D., Ellis, N. C., Miles, T. R., & Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, *11*(2), 185–199. https://doi.org/10.1016/0010-0277(82)90025-7

Barbarotto, R., Capitani, E., Spinnler, H., & Trivelli, C. (1995). Slowly progressive semantic impairment with category specificity. *Neurocase*, *1*(2), 107–119.

https://doi.org/10.1080/13554799508402355

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 560–585. https://doi.org/10.1080/783663595

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654. https://doi.org/10.1037/0278-7393.11.1-4.629

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin and Review*, *23*(4), 1122–1142. https://doi.org/10.3758/s13423-016-1028-3

Barsalou, L. W. (2020). Challenges and opportunities for grounding cognition. *Journal of Cognition*, *3*(1), 1–24. https://doi.org/http://doi.org/10.5334/joc.116

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*(1752), 20170144. https://doi.org/10.1098/rstb.2017.0144

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3), 1–46. https://doi.org/10.1037/h0027577

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin and Review*, *15*(2), 357–363. https://doi.org/10.3758/PBR.15.2.357

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology*, *66*(11), 2135–2160. https://doi.org/10.1080/17470218.2013.775318

Benjafield, J., & Muckenheim, R. (1989). Dates of entry and measures of imagery, concreteness, goodness, and familiarity for 1,046 words sampled from the Oxford English Dictionary. *Behavior Research Methods, Instruments, & Computers*, *21*(1), 31–52. https://doi.org/10.3758/BF03203869

Best, W., Schröder, A., & Herbert, R. (2006). An investigation of a relative impairment in naming non-living items: Theoretical and methodological implications. *Journal of Neurolinguistics*, *19*(2), 96–123. https://doi.org/10.1016/j.jneuroling.2005.09.001

Bi, Y., Han, Z., Shu, H., & Caramazza, A. (2005). Are verbs like inanimate objects? *Brain and Language*, *95*(1), 28–29. https://doi.org/10.1016/j.bandl.2005.07.029

Biedermann, B., Ruh, N., Nickels, L., & Coltheart, M. (2008). Information retrieval in tip of the tongue states: New data and methodological advances. *Journal of Psycholinguistic Research*, *37*(3), 171–198. https://doi.org/10.1007/s10936-007-9065-8

Binney, R. J., Zuckerman, B., & Reilly, J. (2016). A neuropsychological perspective on abstract word representation: from theory to treatment of acquired language disorders. *Current neurology and neuroscience reports*, 16(9), 1-8.

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, and Computers*, *33*(1), 73–79. https://doi.org/10.3758/BF03195349

Bird, H., Howard, D., & Franklin, S. (2000). Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language*, *72*(3), 246–309. https://doi.org/10.1006/brln.2000.2292

Bird, H., Howard, D., & Franklin, S. (2003). Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics*, *16*(2–3), 113–149. https://doi.org/10.1016/S0911-6044(02)00016-7

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*(1), 72–74. https://doi.org/10.1016/S0304-3940(02)00227-6

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*(1), 89–114. https://doi.org/10.1348/000712602162463

Bonin, P., Méot, A., Aubert, L.-F., Malardier, N., Niedenthal, P. M., & Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots [Concreteness, imagery value, subjective frequency and emotional valence norms for 866 words]. *L'année Psychologique*, *103*(4), 655–694. https://doi.org/10.3406/psy.2003.29658

Bonin, P., Méot, A., Ferrand, L., & Roux, S. (2011). L'imageabilité: Normes et relations avec d'autres variables psycholinguistiques [Imageability: Norms and relationships with other psycholinguistic variables]. *L'Année Psychologique*, *111*(2), 327–357. https://doi.org/10.4074/S0003503311002041

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 158–167. https://doi.org/10.3758/BF03195507

Bonin, P., Poulin-Charronnat, B., Lukowski Duplessy, H., Bard, P., Vinter, A., Ferrand, L., & Méot, A. (2020). IMABASE: A new set of 313 colourised line drawings standardised in French for name agreement, image agreement, conceptual familiarity, age-of-acquisition, and imageability. *Quarterly Journal of Experimental Psychology*, *73*(11), 1862–1878. https://doi.org/10.1177/1747021820932822

Bormann, T. (2011). The role of lexical-semantic neighborhood in object naming: Implications for models of lexical access. *Frontiers in Psychology*, *2*, 1–11. 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

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive neuropsychology*, *11*(6), 617-660.

Breedin, S. D., Saffran, E. M., & Schwartz, M. F. (1998). Semantic factors in verb retrieval: An effect of complexity. *Brain and Language*, *63*(1), 1–31. https://doi.org/10.1006/brln.1997.1923

Breining, B. L., Nozari, N., & Rapp, B. (2019). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*(6), 1093–1106. https://doi.org/10.1037/XLM0000630

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*(2), 204–223. https://doi.org/10.1037/0033-2909.109.2.204

Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531–544. https://doi.org/10.3758/BF03196189

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, *114*(104125). https://doi.org/10.1016/j.jml.2020.104125

Butler, R. A., Ralph, M. A. L., & Woollams, A. M. (2014). Capturing multidimensionality in stroke aphasia: Mapping principal behavioural components to neural structures. *Brain*, *137*(12), 3248–3266. https://doi.org/10.1093/brain/awu286

Butterworth, B. (1981). Speech errors: Old data in search of new theories. *Linguistics*, *19*(7–8), 627–662. https://doi.org/10.1515/ling.1981.19.7-8.627

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3–6), 213–261. https://doi.org/10.1080/02643290244000266

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from?. *Cortex*, *26*(1), 95-122.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*(6312), 788–790. https://doi.org/10.1038/349788a0

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*(1), 85–95. https://doi.org/10.1080/14640747308400325

Catricalà, E., della Rosa, P. A., Plebani, V., Perani, D., Garrard, P., & Cappa, S. F. (2015). Semantic feature degradation and naming performance. Evidence from neurodegenerative disorders. *Brain and Language*, *147*, 58–65. https://doi.org/10.1016/j.bandl.2015.05.007

Chen, L., & Rogers, T. T. (2014). Revisiting domain-general accounts of category specificity in mind and brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 327–344. https://doi.org/10.1002/wcs.1283

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From perception to conception: How meaningful objects are processed over time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. https://doi.org/10.1177/014662168300700301

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 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*(3), 452–465. https://doi.org/10.1016/J.COGNITION.2012.07.010

Cosentino, E. (2019). Artifacts and affordances. *Synthese*, *198*(17), 4007–4026. https://doi.org/10.1007/s11229-019-02297-4

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychological Learning, Memory and Cognition*, *32*(4), 643–658. https://doi.org/10.1037/0278-7393.32.4.643

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371–414. https://doi.org/10.1207/s15516709cog2303_4

Crutch, S. J. (2006). Qualitatively different semantic representations for abstract and concrete words: Further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase*, *12*(2), 91–97. https://doi.org/10.1080/13554790500507172

Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: Evidence from odd-one-out judgements. *Quarterly Journal of Experimental Psychology*, *62*(7), 1377–1390. https://doi.org/10.1080/17470210802483834

Crutch, S. J., & Jackson, E. C. (2011). Contrasting graded effects of semantic similarity and association across the concreteness spectrum. *Quarterly Journal of Experimental Psychology (2006)*, *64*(7), 1388–1408. https://doi.org/10.1080/17470218.2010.543285

Crutch, S. J., Ridha, B. H., & Warrington, E. K. (2006). The different frameworks underlying abstract and concrete knowledge: Evidence from a bilingual patient with a semantic refractory access dysphasia. *Neurocase*, *12*(3), 151–163. https://doi.org/10.1080/13554790600598832

Crutch, S. J., & Warrington, E. K. (2003). Preservation of propositional speech in a pure anomic: The importance of an abstract vocabulary. *Neurocase*, *9*(6), 465–481. https://doi.org/10.1076/NEUR.9.6.465.29373

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(3), 615–627. https://doi.org/10.1093/BRAIN/AWH349

Crutch, S. J., & Warrington, E. K. (2007). Semantic priming in deep-phonological dyslexia: Contrasting effects of association and similarity upon abstract and concrete word reading. *Cognitive Neuropsychology*, *24*(6), 583–602. https://doi.org/10.1080/02643290701577351

Crutch, S. J., & Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: Complementary semantic interference and facilitation effects. *Cognitive Neuropsychology*, *27*(1), 46–71. https://doi.org/10.1080/02643294.2010.491359

Cutler, A. (1981a). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*(1–3), 65–70. https://doi.org/10.1016/0010-0277(81)90026-3

Cutler, A. (1981b). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 561–582). De Gruyter Mouton.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195–209. https://doi.org/10.1016/j.jml.2006.11.001

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(2), 345–361. https://doi.org/10.1037/0278-7393.25.2.345

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. https://doi.org/10.3758/s13428-018-1115-7

de Mornay Davies, P., & Funnell, E. (2000). Semantic representation and ease of predication. *Brain and Language*, *73*(1), 92–119. https://doi.org/10.1006/BRLN.2000.2299

de Zubicaray, G., McMahon, K., & Howard, D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in spoken word production. *Language, Cognition and Neuroscience*, *30*(3), 261–272. https://doi.org/10.1080/01690965.2013.848990

Deese, J. (1984). *Thought into speech: The psychology of a language*. Prentice-Hall.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 222–233. https://doi.org/10.1037/0278-7393.10.2.222

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*(4), 588–615. https://doi.org/10.3758/BF03200832

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behavior Research Methods*, *46*(4), 1119–1127. https://doi.org/10.3758/s13428-013-0420-4

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, *10*(1), 77–94. https://doi.org/10.1162/089892998563798

Ellis, A. W., Lum, C., & Lambon Ralph, M. A. (1996). On the use of regression techniques for the analysis of single case aphasic data. *Journal of Neurolinguistics*, *9*(3), 165–174. https://doi.org/10.1016/0911-6044(96)00008-5

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(2), 515–523. https://doi.org/10.1037/0278-7393.24.2.515

Evans, W. S., Hula, W. D., Quique, Y., & Starns, J. J. (2020). How much time do people with aphasia need to respond during picture naming? Estimating optimal response time cutoffs using a multinomial ex-Gaussian approach. *Journal of Speech, Language, and Hearing Research*, *63*(2), 599–614. https://doi.org/10.1044/2019_JSLHR-19-00255

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339–357. https://doi.org/10.1037/0096-3445.120.4.339

Fergadiotis, G., Swiderski, A., & Hula, W. D. (2019). Predicting confrontation naming item difficulty. *Aphasiology*, *33*(6), 689–709. https://doi.org/10.1080/02687038.2018.1495310

Ferré, P., Ventura, D., Comesaña, M., & Fraga, I. (2015). The role of emotionality in the acquisition of new concrete and abstract words. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.00976

Fieder, N., Krajenbrink, T., Foxe, D., Hodges, J. R., Piguet, O., & Nickels, L. (2016). Less is more-Effects of semantic neighbourhood on naming in semantic dementia (svPPA). *Stem-, Spraak- En Taalpathologie*, *21*(Suppl.), 65–68.

Fieder, N., Wartenburger, I., & Abdel Rahman, R. (2019). A close call: Interference from semantic neighbourhood density and similarity in language production. *Memory & Cognition*, *47*(1), 145–168. https://doi.org/10.3758/s13421-018-0856-y

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182–216. https://doi.org/10.1006/jmla.2000.2716

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology*, *12*(5), 549–566. https://doi.org/10.1080/02643299508252007

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27–52. https://doi.org/10.2307/412187

Fromkin, V. A. (Ed.) (1973a). Slips of the tongue. *Scientific American*, *229*(6), 110–117. https://doi.org/10.1038/scientificamerican1273-110

Fromkin, V. A. (1973b). *Speech errors as linguistic evidence*. De Gruyter Mouton.

Funnell, E., & Sheridan, J. (1992). Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology*, *9*(2), 135–153. https://doi.org/10.1080/02643299208252056

Gardner, H. (1973). The contribution of operativity to naming capacity in aphasic patients. *Neuropsychologia*, *11*(2), 213–220. https://doi.org/10.1016/0028-3932(73)90010-9

Garrard, P., Lambon Ralph, M. A., Patterson, K., Pratt, K. H., & Hodges, J. R. (2005). Semantic feature knowledge and picture naming in dementia of Alzheimer's type: A new approach. *Brain and Language*, *93*(1), 79–94. https://doi.org/10.1016/j.bandl.2004.08.003

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of experimental psychology: General*, 113(2), 256. https://doi.org/10.1037/0096-3445.113.2.256

Gianico-Relyea, J. L., & Altarriba, J. (2012). Word concreteness as a moderator of the tip-of-the-tongue effect. *Psychological Record*, *62*(4), 763–776. https://doi.org/10.1007/BF03395834

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Kempler, D., & Seidenberg, M. S. (1997). Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, *57*(2), 254–279. https://doi.org/10.1006/brln.1997.1752

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language*, 24(6), 631-645. https://doi.org/10.1016/0749-596X(85)90050-6

Grober, E., Perecman, E., Kellar, L., & Brown, J. (1980). Lexical knowledge in anterior and posterior aphasics. *Topics in Catalysis*, *10*(2), 318–330. https://doi.org/10.1016/0093-934X(80)90059-0

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, *60*(1), 1–19. https://doi.org/10.1016/j.jml.2008.09.001

Grossman, M. (1981). A bird is a bird is a bird: Making reference within and without superordinate categories. *Brain and Language*, *12*(2), 313–331. https://doi.org/10.1016/0093-934X(81)90022-5

Grossman, M., Robinson, K., Biassou, N., White-Devine, T., & D'Esposito, M. (1998). Semantic memory in Alzheimer's disease: Representativeness, ontologic category, and material. *Neuropsychology*, *12*(1), 34–42. https://doi.org/10.1037/0894-4105.12.1.34

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Gvion, A., & Friedmann, N. (2013). A selective deficit in imageable concepts: A window to the organization of the conceptual system. *Frontiers in Human Neuroscience*, *7*, 1–13. https://doi.org/10.3389/fnhum.2013.00226

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017). Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex*, *86*, 275–289. https://doi.org/10.1016/j.cortex.2016.04.016

Hameau, S., Biedermann, B., Fieder, N., & Nickels, L. (2019). Investigation of the effects of semantic neighbours in aphasia: A facilitated naming study. *Aphasiology*, *34*(7), 840–864. https://doi.org/10.1080/02687038.2019.1652241

Hameau, S., Biedermann, B., Robidoux, S., & Nickels, L. (2021). Effects of phonological neighbourhood density and frequency in picture naming. *Journal of Memory and Language*, *120*(104248), 1–20. https://doi.org/10.1016/j.jml.2021.104248

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

Hart, J., Berndt, R. S., & Caramazza, A. (1985). Category-specific naming deficit following cerebral infarction. *Nature*, *316*(6027), 439–440. https://doi.org/10.1038/316439A0

Hazeltine, E., & Mordkoff, J. T. (2014). Resolved but not forgotten: Stroop conflict dredges up the past. *Frontiers in Psychology*, *5*, 1–8. https://doi.org/10.3389/fpsyg.2014.01327

Hillis, A. E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, *114*(5), 2081–2094. https://doi.org/10.1093/brain/114.5.2081

Hillis, A. E., & Caramazza, A. (1995). The compositionality of lexical semantic representations: Clues from semantic errors in object naming. *Memory*, 3(3-4), 333-358.

Hodges, J. R., Bozeat, S., Lambon Ralph, M. A., Patterson, K., & Spatt, J. (2000). The role of conceptual knowledge in object use: Evidence from semantic dementia. *Brain*, *123*(9), 1913–1925. https://doi.org/10.1093/brain/123.9.1913

Hodgson, C., & Lambon Ralph, M. A. (2008). Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. *Brain and Language*, *104*(1), 89–101. https://doi.org/10.1016/j.bandl.2007.03.007

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition*, *13*(7–8), 884–910. https://doi.org/10.1080/13506280544000093

Hough, M. S., & Pierce, R. S. (1989). Contextual influences on category concept generation in aphasia. In T. Prescott (Ed.), *Clinical Aphasiology* (Vol. 18, pp. 507–520). College-Hill Press. https://doi.org/10.1080/02687038908249022

Howard, D., Best, W., Bruce, C., & Gatehouse, C. (1995). Operativity and animacy effects in aphasic naming. *International Journal of Language & Communication Disorders*, *30*(3), 286–302. https://doi.org/10.3109/13682829509021443

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology*, *20*(9–11), 921–950. https://doi.org/10.1080/02687030600782679

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, *5*(1), 67–104. https://doi.org/10.1080/02643298808252927

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage Publications.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin*, *120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Jones, G. v. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, *24*, 1–19. https://doi.org/10.1016/0093-934X(85)90094-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

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. https://doi.org/10.1016/J.CORTEX.2011.04.006

Kinoshita, S., de Wit, B., & Norris, D. (2017). The magic of words reconsidered: Investigating the automaticity of reading color-neutral words in the Stroop task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(3), 369–384. https://doi.org/10.1037/xlm0000311

Kiran, S., Ntourou, K., & Eubank, M. (2007). The effect of typicality on online category verification of inanimate category exemplars in aphasia. *Aphasiology*, *21*(9), 844–866. https://doi.org/10.1080/02687030600743564

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits. *Journal of Speech, Language, and Hearing Research*, *46*(4), 773–787. https://doi.org/10.1044/1092-4388(2003/061)

Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2007a). Omissions in aphasic picture naming: Late age-of-acquisition is the culprit, not low semantic density. *Brain and Language*, *103*(1–2), 132–133. https://doi.org/10.1016/j.bandl.2007.07.081

Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2007b). Omissions in aphasic picture naming: Late age-of- acquisition is the culprit, as well as low semantic density. *Poster Presented at the 45th Meeting of the Academy of Aphasia, Washington, DC*, October 21–23.

Kohn, S. E., & Smith, K. L. (1995). Serial effects of phonemic planning during word production. *Aphasiology*, *9*(3), 209–222. https://doi.org/10.1080/02687039508248196

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), 14–34. https://doi.org/10.1037/a0021446

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481. https://doi.org/10.1016/J.COGNITION.2009.06.007

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

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400–410. https://doi.org/10.3758/BF03214220

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 613–625. https://doi.org/10.1016/S0022-5371(74)80049-6

Laiacona, M., Barbarotto, R., & Capitani, E. (2006). Human evolution and the brain representation of semantic knowledge: Is there a role for sex differences? *Evolution and Human Behavior*, *27*(2), 158–168. https://doi.org/10.1016/J.EVOLHUMBEHAV.2005.08.002

Laiacona, M., & Capitani, E. (2001). A case of prevailing deficit of nonliving categories or a case of prevailing sparing of living categories? *Cognitive Neuropsychology*, *18*(1), 39–70. https://doi.org/10.1080/02643290042000035

Laiacona, M., Luzzatti, C., Zonca, G., Guarnaschelli, C., & Capitani, E. (2001). Lexical and semantic factors influencing picture naming in aphasia. *Brain and Cognition*, *46*(1–2), 184–187. https://doi.org/10.1016/S0278-2626(01)80061-0

Laisney, M., Giffard, B., Belliard, S., de la Sayette, V., Desgranges, B., & Eustache, F. (2011). When the zebra loses its stripes: Semantic priming in early Alzheimer's disease and semantic dementia. *Cortex*, *47*(1), 35–46. https://doi.org/10.1016/j.cortex.2009.11.001

Lampe, L. F., Fieder, N., Krajenbrink, T., & Nickels, L. (2017). Semantische Nachbarschaft in der Wortproduktion bei Aphasie [Semantic neighbourhood in word production in aphasia]. In A. Adelt, Ö. Yetim, C. Otto, & T. Fritzsche (Eds.), *Spektrum Patholinguistik - Panorama Patholinguistik: Sprachwissenschaft trifft Sprachtherapie* (Vol. 10, pp. 103–114). Universitätsverlag Potsdam. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/39701

Lampe, L. F., Hameau, S., Fieder, N., & Nickels, L. (2021). Effects of semantic variables on word production in aphasia. *Cortex*, *141*, 363–402. https://doi.org/10.1016/j.cortex.2021.02.020

Lampe, L. F., Hameau, S., & Nickels, L. (2022a). Are they really stronger? Comparing effects of semantic variables in speeded deadline and standard picture naming. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218221103356

Lampe, L. F., Hameau, S., & Nickels, L. (2022b). Semantic variables both help and hinder word production: Behavioral evidence from picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 72–97. https://doi.org/10.1037/xlm0001050

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), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Laws, K. R. (2005). “Illusions of normality”: A methodological critique of category-specific naming. *Cortex*, *41*(6), 842–851. https://doi.org/10.1016/S0010-9452(08)70303-4

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122–142. https://doi.org/10.1037/0033-295X.98.1.122

Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F. X. (2014). How familiarization and repetition modulate the picture naming network. *Brain and Language*, *133*, 47–58. https://doi.org/10.1016/j.bandl.2014.03.010

Lloyd-Jones, T. J., & Nettlemill, M. (2007). Sources of error in picture naming under time pressure. *Memory & Cognition*, *35*(4), 816–836. https://doi.org/10.3758/BF03193317

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495. https://doi.org/10.3758/BF03198265

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G. D. (2002). Verb-noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language*, *81*(1–3), 432–444. https://doi.org/10.1006/brln.2001.2536

Lyalka, O., Howard, D., Morris, J., & Nickels, L. (2020). Does producing semantically related words aid word retrieval in people with aphasia? *Aphasiology*, *34*(2), 158–194. https://doi.org/10.1080/02687038.2018.1562149

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *WORD*, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Macoir, J. (2009). Is a plum a memory problem? Longitudinal study of the reversal of concreteness effect in a patient with semantic dementia. *Neuropsychologia*, *47*(2), 518–535. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.10.006

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marshall, J., Pring, T., Chiat, S., & Robson, J. (1996). Calling a salad a federation: An investigation of semantic jargon. Part 1—Nouns. *Journal of Neurolinguistics*, *9*(4), 237–250. https://doi.org/10.1016/S0911-6044(97)82796-0

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, *28*(4), 462–485. https://doi.org/10.1016/0749-596X(89)90022-3

Mason, C., & Nickels, L. (2021). Emotion word processing in aphasia. *Manuscript in Preparation*.

Mason-Baughman, M. B. (2009). *Choosing among related foils in aphasia: The role of common and distinctive semantic features*. [Doctoral dissertation, Kent State University].

Mason-Baughman, M. B. (2010). The role of common and distinctive features in semantic knowledge of persons with aphasia: A pilot study. *PSHA Journal*, 30–40.

Mason-Baughman, M. B., & Wallace, S. E. (2013a). Semantic feature knowledge in persons with aphasia: The role of commonality, distinctiveness, and importance. *Aphasiology*, *27*(3), 364–380. https://doi.org/10.1080/02687038.2012.730602

Mason-Baughman, M. B., & Wallace, S. E. (2013b). The role of commonality, distinctiveness and importance of semantic features in persons with aphasia. *Brain Injury*, *27*(4), 399–407. https://doi.org/10.3109/02699052.2012.750748

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 3–23. https://doi.org/10.1037//0278-7393.21.1.3

Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex*, *45*(6), 738–758. https://doi.org/10.1016/j.cortex.2008.10.003

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

McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, *53*(4), 360–373. https://doi.org/10.1037/h0087323

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130. https://doi.org/10.1037/0096-3445.126.2.99

Melinger, A., & Abdel Rahman, R. (2013). Lexical selection is competitive: Evidence from indirectly activated semantic associates during picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 348–364. https://doi.org/10.1037/a0028941

Meringer, R., & Mayer, C. (1895). *Versprechen und Verlesen: Eine psychologisch-linguistische Studie [To misspeak and misread: A psychological-linguistic study]*. Göschen.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/J.CORTEX.2010.11.002

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations (ICLR)*, 1–12.

Mills, L., Kinoshita, S., & Norris, D. (2019). No negative priming effect in the manual Stroop task. *Frontiers in Psychology*, *10*(1764), 1–9. https://doi.org/10.3389/fpsyg.2019.01764

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex*, *25*(10), 3343–3355. https://doi.org/10.1093/cercor/bhu137

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cognitive, Affective and Behavioral Neuroscience*, *11*(1), 32–43. https://doi.org/10.3758/s13415-010-0009-7

Mirman, D., & Graziano, K. M. (2013). The neural basis of inhibitory effects of semantic and phonological neighbors in spoken word production. *Journal of Cognitive Neuroscience*, *25*(9), 1504–1516. https://doi.org/10.1162/jocn_a_00408

Mirman, D., Kittredge, A. K., & Dell, G. S. (2010). Effects of near and distant phonological neighbors on picture naming. *Proceedings of the Annual Meeting of the Cognitive Science Society, 32*, 1447–1452.

Mirman, D., & Magnuson, J. S. (2006). The impact of semantic neighborhood density on semantic access measures of semantic neighborhood density. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 1823–1828.

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*(1), 65–79. https://doi.org/10.1037/0278-7393.34.1.65

Mirman, D., Strauss, T. J., Brecher, A., Walker, G. M., Sobel, P., Dell, G. S., & Schwartz, M. F. (2010). A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function. *Cognitive Neuropsychology*, *27*(6), 495–504. https://doi.org/10.1080/02643294.2011.574112

Mirman, D., Zhang, Y., Wang, Z., Coslett, H. B., & Schwartz, M. F. (2015). The ins and outs of meaning: Behavioral and neuroanatomical dissociation of semantically-driven word retrieval and multimodal semantic recognition in aphasia. *Neuropsychologia*, *76*, 208–219. https://doi.org/10.1016/j.neuropsychologia.2015.02.014

Morelli, C. A., Altmann, L. J. P., Kendall, D., Fischler, I., & Heilman, K. M. (2011). Effects of semantic elaboration and typicality on picture naming in Alzheimer disease. *Journal of Communication Disorders*, *44*(4), 413–428. https://doi.org/10.1016/J.JCOMDIS.2011.01.006

Morrison, C. M. (2003). Interpret with caution: Multicollinearity in multiple regression of cognitive data. *Perceptual and Motor Skills*, *97*(1), 80–82. https://doi.org/10.2466/pms.2003.97.1.80

Moses, M. S., Nickels, L., & Sheard, C. (2004). “I’m sitting here feeling aphasic!” A study of recurrent perseverative errors elicited in unimpaired speakers. *Brain and Language*, *89*(1), 157–173. https://doi.org/10.1016/S0093-934X(03)00364-X

Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). ‘Two eyes of a see-through’: Impaired and intact semantic knowledge in a case of selective deficit for living things. *Neurocase*, *4*(4–5), 291–310. https://doi.org/10.1093/neucas/4.4.291

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

Newcombe, F. B., Oldfield, R. C., & Wingfield, A. (1965). Object-naming by dysphasic patients. *Nature*, *207*(5002), 1217–1218. https://doi.org/10.1038/2071217a0

Newton, C., Thornley, H., & Bruce, C. (2020). The influence of emotional valence on word recognition in people with aphasia. *Language, Cognition and Neuroscience*, *35*(8), 1064–1072.

https://doi.org/10.1080/23273798.2020.1713385

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. *Language and Cognitive Processes*, *10*(1), 13–45. https://doi.org/10.1080/01690969508407086

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Psychology Press.

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology*, *11*(3), 289–320. https://doi.org/10.1080/02643299408251977

Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*(10), 1281–1303. https://doi.org/10.1016/0028-3932(95)00102-9

Nickels, L., Howard, D., & Best, W. (2011). On the use of different methodologies in cognitive neuropsychology: Drink deep and from several sources. *Cognitive Neuropsychology*, 28(7), 475-485.

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior Research Methods*, *44*(4), 934–945. https://doi.org/10.3758/s13428-011-0176-7

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*(3), 434–455. https://doi.org/10.1037/a0012799

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, *53*, 1–19. https://doi.org/10.1016/j.jneuroling.2019.100875

Obermeyer, J., Reinert, L., Kamen, R., Pritchard, D., Park, H., & Martin, N. (2022). Effect of working memory load and typicality on semantic processing in aphasia. *American Journal of Speech-Language Pathology*, *31*(1), 12–29. https://doi.org/10.1044/2021_AJSLP-20-00283

Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, *202*(4936), 1031–1032. https://doi.org/10.1038/2021031a0

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. https://doi.org/10.1080/17470216508416445

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114(2), 227–252.

Oppenheim, G., & Nozari, N. (2021). Behavioral interference or facilitation does not distinguish between competitive and noncompetitive accounts of lexical selection in word production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43). https://escholarship.org/uc/item/48z6c6gk

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. https://doi.org/10.1037/H0027272

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Erlbaum.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon*, *5*(2), 205–230. https://doi.org/10.1075/ml.5.2.04pai

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, *142*(1), 282–287. https://doi.org/10.1037/a0027004

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1, Pt.2), 1–25. https://doi.org/10.1037/h0025327

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

Perret, C., & Laganaro, M. (2013). Why are written picture naming latencies (not) longer than spoken naming? *Reading and Writing*, *26*(2), 225–239. https://doi.org/10.1007/s11145-012-9365-8

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Meaning and context, individuals and development* (pp. 24–43). Psychology Press.

Pexman, P. M. (2020). How does meaning come to mind? Four broad principles of semantic processing. *Canadian Journal of Experimental Psychology*, *74*(4), 275–283. https://doi.org/10.1037/CEP0000235

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

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. https://doi.org/10.3758/s13428-018-1171-z

Plaut, D. C. (1996). Relearning after Damage in Connectionist Networks: Toward a Theory of Rehabilitation. *Brain and Language*, *52*(1), 25–82. https://doi.org/10.1006/brln.1996.0004

Plaut, D. C., & Shallice, T. (1991). Effects of word abstractness in a connectionist model of deep dyslexia. *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, 73–78.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*(5), 377–500. https://doi.org/10.1080/02643299308253469

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, *21*(e12549), 1–12. https://doi.org/10.1111/desc.12549

Ponari, M., Norbury, C. F., & Vigliocco, G. (2020). The role of emotional valence in learning novel abstract concepts. *Developmental Psychology*, *56*(10), 1855–1865. https://doi.org/10.1037/dev0001091

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*(1), 68–89. https://doi.org/10.1016/j.cognition.2014.03.010

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240–244. https://doi.org/10.1016/j.cognition.2015.09.016

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2021). Semantic richness and density effects on language production: Electrophysiological and behavioral evidence. *Journal of Experimental Psychology: Learning Memory and Cognition*, *47*(3), 508–517. https://doi.org/10.1037/xlm0000940

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 393–406. https://doi.org/10.1037/0278-7393.30.2.393

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460. https://doi.org/10.1037/0033-295x.107.3.460

Recchia, G., & Jones, M. N. (2012). The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, *6*, 1–16. https://doi.org/10.3389/fnhum.2012.00315

Reilly, J., Peelle, J. E., & Grossman, M. (2007). A unitary semantics account of reverse concreteness effects in semantic dementia. *Brain and Language*, *1*(103), 86-87.

Rico Duarte, L., Marquié, L., Marquié, J. C., Terrier, P., & Ousset, P. J. (2009). Analyzing feature distinctiveness in the processing of living and non-living concepts in Alzheimer's disease. *Brain and Cognition*, *71*(2), 108–117. https://doi.org/10.1016/j.bandc.2009.04.007

Rico Duarte, L., & Robert, C. (2014). Semantic richness influences naming pictures in old age and in Alzheimer's disease. *Journal of Neurolinguistics*, *31*, 55–68. https://doi.org/10.1016/j.jneuroling.2014.06.004

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, *4*(2), 131–185. https://doi.org/10.1080/02643298708252038

Riley, E. A., & Thompson, C. K. (2010). Semantic typicality effects in acquired dyslexia: Evidence for semantic impairment in deep dyslexia. *Aphasiology*, *24*(6–8), 802–813. https://doi.org/10.1080/02687030903422486

Rofes, A., Zakariás, L., Ceder, K., Lind, M., Johansson, M. B., de Aguiar, V., Bjekić, J., Fyndanis, V., Gavarró, A., Simonsen, H. G., Sacristán, C. H., Kambanaros, M., Kraljević, J. K., Martínez-Ferreiro, S., Mavis, İ., Orellana, C. M., Sör, I., Lukács, Á., Tunçer, M., … Howard, D. (2018). Imageability ratings across languages. *Behavior Research Methods*, *50*(3), 1187–1197. https://doi.org/10.3758/s13428-017-0936-0

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220–239. https://doi.org/10.1016/j.neuropsychologia.2015.04.015

Rogerson, P. (2011). *Statistical methods for geography*. SAGE Publications, Ltd. https://doi.org/10.4135/9781849209953

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rose, S. B., & Abdel Rahman, R. (2017). Semantic similarity promotes interference in the continuous naming paradigm: Behavioural and electrophysiological evidence. *Language, Cognition and Neuroscience*, *32*(1), 55–68. https://doi.org/10.1080/23273798.2016.1212081

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 753–763. https://doi.org/10.1037/xlm0000592

Rossiter, C., & Best, W. (2013). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, *27*(7), 784–798. https://doi.org/10.1080/02687038.2012.751579

Sandberg, C., Sebastian, R., & Kiran, S. (2012). Typicality mediates performance during category verification in both ad-hoc and well-defined categories. *Journal of Communication Disorders*, *45*(2), 69–83. https://doi.org/10.1016/j.jcomdis.2011.12.004

Sartori, G., & Job, R. (1988). The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology*, *5*(1), 105–132. https://doi.org/10.1080/02643298808252928

Sartori, G., & Lombardi, L. (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, *16*(3), 439–452. https://doi.org/10.1162/089892904322926773

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press. https://doi.org/10.1017/CBO9780511526817

Shallice, T. (1991). Précis of From neuropsychology to mental structure. *Behavioral and Brain Sciences*, *14*(3), 429–438. https://doi.org/10.1017/S0140525X0007059X

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109–136). Springer.

https://doi.org/10.1007/978-1-4613-8202-7_6

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17–51). Lawrence Erlbaum Associates, Inc.

Sirigu, A., Duhamel, J. R., & Poncet, M. (1991). The role of sensorimotor experience in object recognition: A case of multimodal agnosia. *Brain*, *114*(6), 2555–2573. https://doi.org/10.1093/brain/114.6.2555

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes*, *28*(5), 597–614. https://doi.org/10.1080/01690965.2012.718088

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605. https://doi.org/10.3758/BF03193891

Stanczak, L., Waters, G., & Caplan, D. (2006). Typicality-based learning and generalisation in aphasia: Two case studies of anomia treatment. *Aphasiology*, *20*(2–4), 374–383. https://doi.org/10.1080/02687030600587631

Starreveld, P. A., & La Heij, W. (2017). Picture-word interference is a Stroop effect: A theoretical analysis and new empirical findings. *Psychonomic Bulletin and Review*, *24*(3), 721–733. https://doi.org/10.3758/s13423-016-1167-6

Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, *56*(3–4), 235–259. https://doi.org/10.1016/0024-3841(82)90012-2

Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data. In B. J. Baars (Ed.), *Experimental slips and human error* (pp. 195–215). Springer. https://doi.org/10.1007/978-1-4899-1164-3_8

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, *25*(2), 163–180. https://doi.org/10.1016/0749-596X(86)90027-6

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Taylor, K. I., Devereux, B. J., Acres, K., Randall, B., & Tyler, L. K. (2012). Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition*, *122*(3), 363–374. https://doi.org/10.1016/j.cognition.2011.11.001

Taylor, K. I., Moss, H., Randall, B., & Tyler, L. K. (2004). The interplay between distinctiveness and intercorrelation in the automatic activation of word meaning. *Proceedings of the 45th Annual Meeting of the Psychonomic Society*, *9*, 109.

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

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*(6), 244–252. https://doi.org/10.1016/S1364-6613(00)01651-X

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language*, *75*(2), 195–231. https://doi.org/10.1006/brln.2000.2353

Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: A trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in Neuroscience*, *8*, 1–13. https://doi.org/10.3389/fnins.2014.00390

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*, *6*(2), 227. https://doi.org/10.4172/2161-1165.1000227

Vieth, H. E., McMahon, K. L., & de Zubicaray, G. I. (2014). The roles of shared vs. distinctive conceptual features in lexical access. *Frontiers in Psychology*, *5*, 1–12. https://doi.org/10.3389/fpsyg.2014.01014

Vigliocco, G., & Vinson, D. P. (2007). Semantic representation. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 195–214). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.001.0001

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/J.NEUBIOREV.2010.04.007

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*(4), 422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition*, *11*(3), 403–430. https://doi.org/10.1017/LANGCOG.2019.23

Vinson, D. P., Andrews, M., & Vigliocco, G. (2013). Giving words meaning. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.) *The Oxford handbook of language production*, Chapter 8, p.134–151 Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.006

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. https://doi.org/10.3758/BRM.40.1.183

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747. https://doi.org/10.1037/0278-7393.28.4.735

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 664–680. https://doi.org/10.1037/0278-7393.17.4.664

Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 243–259. https://doi.org/10.1037/0278-7393.19.2.243

Wallace, S. E., & Mason-Baughman, M. B. (2012). Relationship between distinctive feature knowledge and word retrieval abilities in people with aphasia. *Aphasiology*, *26*(10), 1278–1297. https://doi.org/10.1080/02687038.2012.702886

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

Warrington, E. K. (1981). Concrete word dyslexia. *British Journal of Psychology, 72*(2), 175–196. https://doi.org/10.1111/j.2044-8295.1981.tb02175.x

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853. https://doi.org/10.1093/brain/107.3.829

Woollams, A. M. (2012). Apples are not the only fruit: The effects of concept typicality on semantic representation in the anterior temporal lobe. *Frontiers in Human Neuroscience*, *6*, 1–9. https://doi.org/10.3389/fnhum.2012.00085

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia*, *46*(10), 2503–2514. https://doi.org/10.1016/j.neuropsychologia.2008.04.005

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition.

Frontiers in Human Neuroscience, *6*, 1–10. https://doi.org/10.3389/fnhum.2012.00072

Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin and Review*, *21*(2), 526–533. https://doi.org/10.3758/s13423-013-0525-x