



HAL
open science

Word-association norms for 1,100 French words with varying levels of concreteness

Dounia Lakhzoum, Marie Izaute, Ludovic Ferrand

► **To cite this version:**

Dounia Lakhzoum, Marie Izaute, Ludovic Ferrand. Word-association norms for 1,100 French words with varying levels of concreteness. *Quarterly Journal of Experimental Psychology Series a Human Experimental Psychology*, 2023, 76 (12), pp.2794-2803. 10.1177/17470218231154454 . hal-03920422

HAL Id: hal-03920422

<https://uca.hal.science/hal-03920422>

Submitted on 21 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Word association norms for 1 100 French words with varying levels of concreteness

Dounia Lakhzoum, Marie Izaute, and Ludovic Ferrand

Université Clermont Auvergne, CNRS, LAPSCO, F-63001 Clermont-Ferrand, France

Quarterly Journal of Experimental Psychology (2023).

<https://doi.org/10.1177/17470218231154454>

Corresponding author: Dounia Lakhzoum or Ludovic Ferrand, Université Clermont Auvergne, CNRS LAPSCO, 34 avenue Carnot TSA 60401- 63001 Clermont-Ferrand, FRANCE. Email: dounia.lakhzoum@gmail.com; ludovic.ferrand@uca.fr

Author's note

This research was financed by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25).

ACKNOWLEDGMENT

This work was sponsored by a public grant overseen by the French National Research Agency as part of the “Investissements d’Avenir” through the IDEX-ISITE initiative CAP 20-25 (ANR-16-IDEX-0001).

ABSTRACT

The organization of concepts in the mental lexicon is a widely studied research topic in experimental psychology. For instance, several studies have shown that whereas concrete concepts are organised according to semantic similarity, abstract concepts are organised according to verbal association. However, these results are not systematically replicated, mainly due to a lack of normative database especially in French. To that end, we introduce a French word association database for 1 100 cues with varying levels of concreteness from abstract to concrete concepts. Analyses from the word association task revealed stronger association strength for concrete concepts compared to abstract concepts. Additional results showed that cues tend to elicit responses of a similar level of concreteness. The database will be useful for investigators interested in French verbal associations for abstract and concrete concepts. The data (available on OSF <https://osf.io/dhuqs/>) introduce responses organised according to association strength and provides cue concreteness.

Keywords: Word association norms, Concreteness, Normative ratings, French words.

INTRODUCTION

Psycholinguistic experiments which study the organisation of concepts in the mental lexicon often rely on normative stimuli. Although there are numerous types of relationships between concepts from lexical and distributional to sensorimotor components (Günther et al., 2022; Landauer & Dumais, 1997; Lund & Burgess, 1996; Mikolov et al., 2013; Villani, Lugli, Liuzza, Nicoletti & Borghi, 2021; Wingfield & Connell, 2022a; 2022b), two types of norms have a long tradition in the field: semantic similarity and word association norms. Semantic similarity refers to the overlap between the defining features of concepts (for example <bee> and <wasp> are semantically similar). In the case of word associations, two words are considered associates when they frequently occur together across language (for example <bee> and <honey> are verbal associates; Ferrand & Alario, 1998; McNamara, 1992; Plaut, 1995).

Semantic and associative priming

The development of normative databases addresses the crucial need to access carefully crafted material when studying cognitive phenomena such as free recall (Nelson, McEvoy & Schreiber, 2004), age-related lexical access (Taconnat et al., 2009), metamemory (Bugaiska, Morson, Moulin & Souchay, 2011; Eakin, 2005; West & Mulligan, 2019), or clinical pathologies such as schizophrenia (Kircher, Whitney, Krings, Huber & Weis, 2008) and semantic refractory access dysphasia (Crutch & Warrington, 2005).

In psycholinguistics, semantic and associative priming effects rely on semantic and associative relationships respectively with some overlap between the two (for example <cat> and <dog> are semantically similar but also verbal associates; Ferrand & New, 2003; Hutchison, 2003; McRae & Boisvert, 1998; Perea & Rosa, 2002; Thompson-Schill, Kurtz, & Gabrieli, 1998). Previous studies have shown that an overlap between semantic and associative relationships could create some discrepancies in reported results (see, for example,

Hutchison, 2003 but also Thompson-Schill et al., 1998). McRae and Boisvert (1998) suggested we use semantically similar pairs that are not associates (for example <whale> and <dolphin>). Ferrand and New (2003) obtained a pure semantic priming effect by using stimuli that were semantically similar while controlling for verbal association (e.g., <dolphin> and <whale>) and a pure associative priming effect by using strongly associated stimuli with no semantic similarity (e.g., <spider> and <web>). Consequently, researchers have used word association norms either to select associated word pairs or to check that semantically similar pairs are not also verbally associated.

Recent studies have shown that concept relationships are central to linguistic models as well. Hill and colleagues (2015) consider that “association and similarity are neither mutually exclusive nor independent” (Hill, Reichart & Korhonen, 2015; p. 668). For instance, the concepts <bike> and <car> are related by both relations to some degree while <car> and <petrol> are strongly verbally associated but not semantically similar (Hill et al., 2015).

De Deyne, Navarro, Perfors, Brysbaert and Storms (2019) qualified this propensity to consider verbal association as a nuisance-variable that must be controlled when manipulating semantic similarity between concepts as a limited approach and suggested that associative data themselves can provide a strong indication of similarity.

In addition, recent models of corpus-based linguistic distributional models (LDMs) which do not discriminate between semantic and associative relationships have been introduced as cognitively plausible models able to capture human performance in various linguistic tasks (De Deyne et al., 2019; Schloss & Li, 2016; Wingfield & Connell, 2022a; but see De Deyne, Verheyen & Storms, 2016). Such modelling approaches assume that linguistic distributional knowledge when completed with sensorimotor components can represent essential aspects of semantics (Wingfield & Connell, 2022a).

Concreteness and organisation of the mental lexicon

There is no clear-cut dichotomy between concrete and abstract concepts (Della Rosa et al., 2010). The Dual Coding Theory (Paivio, Yuille & Madigan, 1968) suggests that concrete concepts are processed based on both a visual and a verbal system, while abstract concepts are processed only in the verbal system. The Context Availability Theory (Schwanenflugel, Harnishfeger & Stowe, 1988) posits that while concrete concepts refer to a definite number of contexts, abstract concepts are connected to varied contexts. More recently, embodied and grounded approaches have defined abstract concepts as intangible entities grounded in emotions, events, social and introspective dimensions (see Borghi et al., 2017 for a review).

The importance of the concreteness variable is further borne out by the development of several widely used concreteness rating norms (see Brysbaert, Warriner & Kuperman, 2014; Coltheart, 1981 in English; Bonin, Méot & Bugajska, 2018 in French). Recent studies have shown that concreteness must be taken into account in addition to association strength and semantic similarity in order to have a fuller picture of how the mental lexicon is organised. Crutch and colleagues (2005; Crutch, Connell & Warrington, 2009; Crutch & Warrington, 2010) reported a case study of a patient with symptoms of semantic refractory access dysphasia. They found that concrete concepts elicited only a semantic priming effect and no associative priming effect, whereas abstract concepts elicited the opposite pattern. These findings suggest that while concrete concepts are organised according to semantic similarity, abstract concepts are organised according to verbal association. Several studies tried to replicate this pattern with some discrepancies (e.g., Duñabeitia, Avilés, Afonso, Scheepers & Carreiras, 2009; Geng & Schnur, 2015; Hamilton & Coslett, 2008). Crutch and Jackson (2011) suggested the disparities might be due to how the concreteness variable was operationalised. In two semantic odd-one-out judgement and matching-to-sample paradigms,

they used triplets of low, middle and high levels of concreteness. Results showed that the semantic effect increased with concreteness while the associative effect decreased.

Furthermore, they suggested that concreteness be used as a graded variable rather than a binary one, especially when studying its effect on the organisation of semantic representation.

Accounts of semantic representation

In recent years, two main accounts of the nature of semantic representation have been debated, namely the distributional account (Firth, 1957; Harris, 1954; see Andrews, Frank & Vigliocco, 2014; Andrews, Vigliocco & Vinson, 2009; Bruni, Tran & Baroni, 2014; Lenci, 2018 for reviews of the distributional account) and embodiment account (see Barsalou, 1999; Glenberg, 1997; Meteyard, Cuadraro, Bahrami & Vigliocco, 2012; Ostarek & Huettig, 2019; Pulvermüller, 2013; Zwaan, 2004; for reviews of the embodied account). They differ in terms of the information used to represent meaning. While distributional semantics relies on symbolic and linguistic features, embodiment relies on modality-specific information.

According to the distributional hypothesis, words occurring in similar contexts have similar meanings (Harris, 1954). The meaning of both abstract and concrete concepts can therefore be derived from statistical co-occurrences across large corpora. Hence, this account makes great use of variables, such as semantic and relatedness variables, that describe associations between words (see Andrews, Frank & Vigliocco, 2014; Borghi et al., 2017; Louwerse & Jeuniaux, 2010). It is an account which has come under attack in recent years, however, because according to embodied theories it fails to explain how concepts are grounded in world referents (see Barsalou, 2020; Borghi et al., 2017 for reviews).

The embodied account, on the other hand, defines semantic representation as being grounded in perceptual and motor states (Barsalou, 1999; Glenberg, 1997; Meteyard, Cuadrado, Bahrami & Vigliocco, 2012; Kiefer & Pulvermüller, 2012; Zwaan, 2004). Studies

have shown evidence of behavioural and brain activation of modality-specific processes, which makes it better suited to explain the processing of concrete concepts over abstract concepts (Meteyard, et al., 2012; Kiefer & Pulvermüller, 2012). Recent hypotheses have been formulated to account for the processing of abstract concepts within the embodied framework, with studies showing that abstract concepts are grounded in social and introspective dimensions (Barsalou & Wiemer-Hastings, 2005), emotional features (Kousta, Vigliocco, Vinson, Andrews & Del Campo, 2011; Lenci, Lebani & Passaro, 2018), information about events, and thematic roles (Ferretti, McRae & Hatherell, 2001).

This debate has evolved in recent years¹ due to the introduction of models capable of capturing real-world experience via linguistic associations (see Lynott et al., 2020; Wingfield & Connell, 2022b for example). This is possible because there are statistical redundancies between the structure of the perceivable world and language so relations between words tend to reproduce the relations between their referents in the real world (Johns & Jones, 2012).

To capture relations other than word associations for both abstract and concrete concepts new variables have emerged. For example, the perceptual strength variable developed by Chedid, Brambati, Bedetti, Rey, Wilson and Vallet (2019; see also Lynott, Connell, Brysbaert, Brand & Carney, 2020) aim to capture perceptual experiences. It represents the extent to which a concept relies on sensorimotor traits for its representation. It contributes to the study of relations between word associations and perceptual modalities.

Word associations and normative datasets

In a word association task, participants are presented with a word (the cue) and asked to generate the first word that comes to mind (the response). The task is particularly effective

¹ Also as suggested by one reviewer who we would like to thank for their insight.

and powerful because it demands barely any cognitive cost and is easily gamified, and yet it can unveil shared mental representations, language patterns and mechanisms of meaning extraction (De Deyne & Storms, 2008; De Deyne et al., 2019; Mollin, 2009; Prior & Bentin, 2008). Word association norms have mostly been collected in English. The *Edinburgh Associative Thesaurus* (Kiss, Armstrong, Milroy & Piper, 1972), with more than 8 000 cues, is one of the most commonly used. More recently, Moss and Older (1996) introduced the *Birkbeck norms* with over 2 600 cues in British English. The more recent set of *University of South Florida norms* (Nelson et al., 2004) introduced free association data for 72 000 word pairs and is the most widely used, with more than 2 000 citations from psycholinguistics to computational linguistics. Even more recently, De Deyne et al., (2019) launched the *Small World of Words* project (*SWOW-EN*), an ongoing megastudy which aims to capture most word associations. It consists of over 12 000 cue words, each of which has elicited 300 responses from half a million participants at the time of writing this paper, making this dataset the most comprehensive one to date. This particular project followed a similar one by De Deyne, Navarro and Storms (2013) which consisted of 16 000 cues in Dutch (*SWOW-NL*). Other languages also have word association datasets. The number of cues is still significant, albeit reduced, and the datasets are still highly valuable for researchers studying these languages, such as Korean (Jung, Na & Akama, 2010), Japanese (Joyce, 2005) and Chinese (Kwong, 2013). In Spanish, Fernández, Díez, Alonso and Beato (2004) introduced free-association norms that is constantly updated with more than 6,000 words at the time of writing, and finally, Barrón-Martínez and Arias-Trejo (2014) presented the *Spanish Word Association Norms* for 234 concrete nouns in Mexican Spanish. In French, Ferrand and Alario (1998) introduced word associations for 366 concrete nouns. They were quickly followed by another dataset of 260 abstract words (Ferrand, 2001). Later on, Duscherer and Mounoud (2006) collected free association data for 151 French action verbs, and, more recently, Bonin,

Méot, Ferrand and Bugaïška (2013) collected word associations for 520 concrete words. As evidenced by all these datasets, the history of the collection of word associations in French is rather spread out in time, making the present study the first for nearly a decade and the one involving the largest number of cues from the largest number of participants for the French language.

The present study

To address this gap in French norms, the present study introduces a dataset comprising word association norms for 1 100 cues obtained from French participants. The present study is a continuation of our previous study which introduced 630 French semantically similar pairs (Lakhzoum et al., 2021). These pairs were composed of a cue and semantically similar related target. The materials included in the present study is composed of the word pairs that were separated in isolated words and presented to participants as cue words to which participants produced associated response words.

The deliberate effort to collect responses for cues of varying levels of concreteness represents the main strength for these norms compared to previous norms. Given that few norms have introduced varying levels of concreteness, we compared concrete and abstract cues with respect to the lexical frequency and the associative strength of the associated words produced. We added a measure of cue word concreteness collected in a previous study (Lakhzoum et al., 2021) which will allow experimenters to select cues with varying levels of concreteness.

METHOD

Participants

The word association task was presented as an online questionnaire. All participants were native French-speakers aged between 18 and 45 years old. We collected data from 1 200 participants (627 women; $M_{age} = 29.30$; $SD = 7.48$). Participants volunteered in response to an announcement posted on Facebook group walls. All participants that were older than 45 years old or non-native speakers were filtered during the recruitment process. The data from 2 participants were invalid and had to be removed. The following analyses are based on the data from 1198 participants. The study was approved by the Research Ethics Committee of Université Clermont Auvergne.

Stimuli

The stimuli were composed of the isolated words used to create the 630 word-pairs from Lakhzoum, Izaute and Ferrand (2021). To ensure the level of concreteness, the words were selected from Coltheart's (1981) concreteness norms (range 100-600). Figure 1 presents a distribution of the cue levels of concreteness. The selected words were then translated into French following a back-translation procedure (Sperber, Devellis & Boehlecke, 1994). For the present study, the pairs were dissociated. After removal of repeated words, a total of 1100 cue words were kept (898 nouns, 106 adjective, 93 verbs and 3 adverbs). Cues and targets were separated in different lists so that participants would not see two words that shared semantic features to avoid influencing their responses and the order of the cues was randomised. To that end, and to keep the task short, since participants tend to be less focused in online tasks than lab tasks, we divided the material into 11 lists of 100 words each.

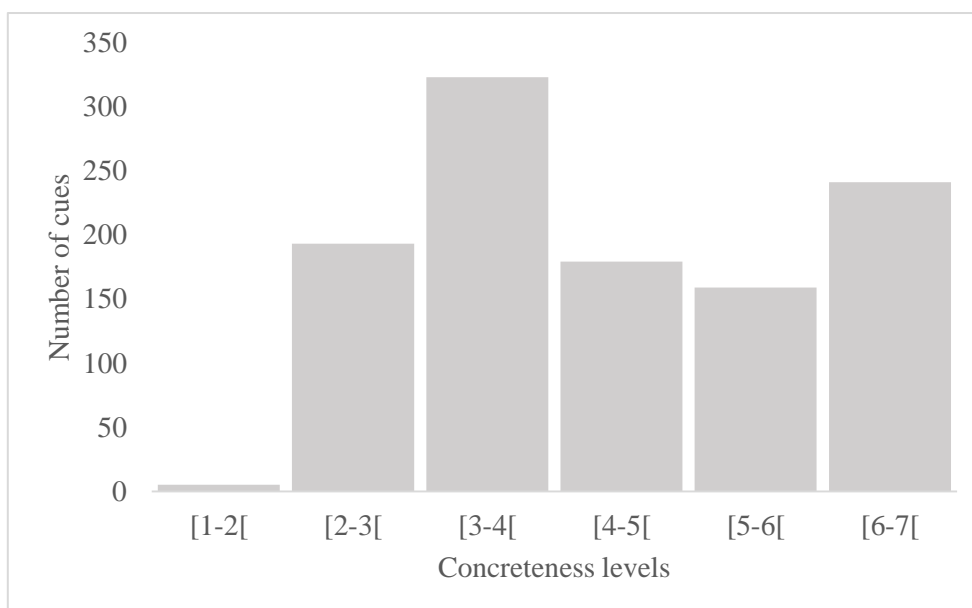


Figure 1. Distribution of cue words according to concreteness levels (1 = very abstract; 7 = very concrete).

Cue Descriptive statistics

General statistics were first computed for the cue words. The mean concreteness variable reported was taken from the Lakhzoum et al. (2021) concreteness study. The general statistics for the lexical variables presented in Table 1 were obtained by cross-referencing the present cue words with the Lexique (New et al., 2004), FLP (Ferrand et al., 2010), MEGALEX (Ferrand et al., 2018) and Wordlex (Gimenes & New, 2016) databases. The concreteness statistics indicate that the present dataset contains cue words of varying levels of concreteness ranging from 1.79 to 6.94 (median = 4.18) on a 7-point scale allowing users to study responses elicited by both abstract and concrete concepts.

Table 1. Descriptive and behavioural data for cue words

| | <i>min</i> | <i>max</i> | <i>M</i> | <i>SD</i> |
|--|------------|------------|----------|-----------|
| Mean Concreteness ^a | 1.79 | 6.94 | 4.45 | 1.41 |
| Movie Frequency ^b | 0.00 | 2752.00 | 440.15 | 136.47 |
| Book Frequency ^b | 0.00 | 4696.15 | 44.92 | 172.65 |
| Blog Frequency ^c | 0.06 | 3095.93 | 61.24 | 193.06 |
| Twitter Frequency ^c | 0.00 | 4070.02 | 55.45 | 232.49 |
| Newspaper Frequency ^c | 0.00 | 3782.26 | 57.49 | 200.59 |
| N-letters ^b | 3.00 | 14.00 | 7.00 | 2.02 |
| N-orthographic Neighbours ^b | 0.00 | 23.00 | 2.42 | 3.60 |
| OLD20 ^d | 1.00 | 5.40 | 2.09 | 0.62 |
| Reaction Times (FLP) ^e | 515.54 | 1005.15 | 658.41 | 65.67 |
| Reaction Times (Megalex) ^f | 473.48 | 747.84 | 553.05 | 44.18 |

Note. a: Lakhzoum et al. (2021); b: New et al. (2004); c: Gimenes & New (2016); d: OLD20: orthographic Levenshtein distance (Yarkoni, Balota, & Yap, 2008; Ferrand et al., 2010); e: FLP: French Lexicon Project (Ferrand et al., 2010); f: Ferrand et al. (2018).

Procedure

The stimuli were divided into 11 lists of 100 isolated cues. Each participant was randomly assigned to only one of the lists. The experiment was conducted online using Qualtrics software (2020). The experiment interface was designed to allow participants to complete the task on either a computer or smartphone. Their task was to read the word that appeared on the screen and to answer in the space below with the first word that came to mind as quickly as possible. The cue words appeared on the screen one after another. They were replaced as soon as the participant had answered. To avoid unacceptable responses, when participants could not think of an answer, they were given the opportunity of leaving a blank and moving onto the next cue. The cues were presented in the middle of the screen in Arial 12 font against a white background. No training was given before the tasks started. The study was self-paced, with no time limit for either the stimulus presentation or the participant's answer. The task took about 15 minutes to complete.

RESULTS

Data Pre-processing

Before performing any analysis, the dataset was cleaned to remove any responses that failed to meet our criteria for inclusion. Symbols and punctuation such as tags, quotes, final punctuation and double spaces were removed. We also removed responses that seemed to indicate participants did not know what to answer. For instance, they sometimes entered “rien”, “aucune réponse” or “aucun mot” in French (*nothing, no response and no word*), indicating that no response came to mind, instead of leaving a blank as the procedure allowed them to do. Finally, responses that were phrases intended to describe what the cue brought to mind were removed. This concerned 523 responses. The responses were then spell-checked and lemmatized by cross-referencing the dataset with Lexique (New et al., 2004). By the end of the process, 2.5% of the data had been removed. Analyses were performed on the remaining 89 707 unique responses or *tokens*.

Distributional statistics

A first exploration of the dataset focused on its distributional characteristics. When all responses were aggregated, there were 10 604 distinct words or *types*, 4 233 of which appeared only once. The responses occurring only once are known as *hapax legomena* responses. In accordance with De Deyne et al. (2013, 2018, but see Nelson et al., 2004), we decided to keep these responses because they reflect the long tail of the frequency spectrum. When sorting the responses according to cue median concreteness, the hapax legomena were equally distributed at 53% for the responses elicited by abstract cues and 55% for the responses elicited by concrete cues. These aspects of quantitative linguistics can be a good tool for capturing common linguistic phenomena across datasets of varying sizes. For

instance, Figure 1 is a frequency spectrum of the distribution of the number of types (V_m) as a function of frequency class (m).

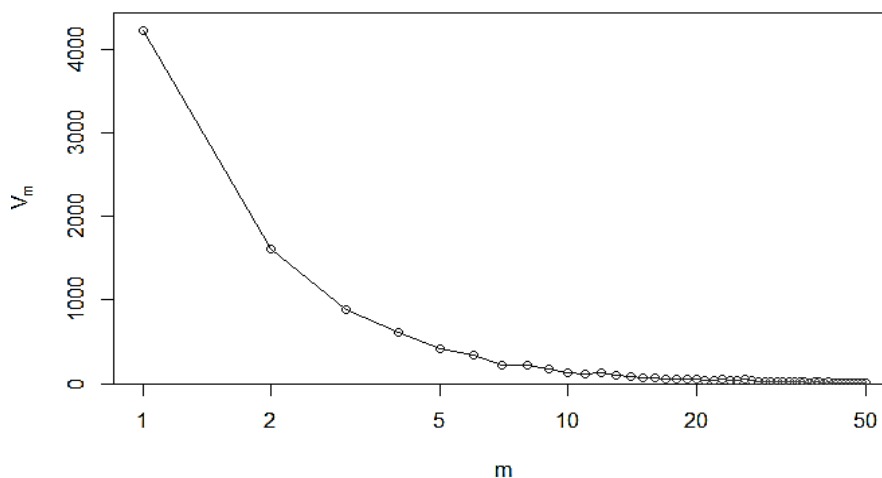


Figure 2. Frequency spectrum on a logarithmic scale representing the number of types (V_m) as a function of frequency class (m). The class corresponding to $m = 1$ corresponds to the hapax legomena.

Figure 2 shows that there are more than 4 000 words in frequency class V_1 , which represents the *hapax legomena*, but very few words in frequency class V_{50} . This pattern, characterized by few exemplars in low frequency classes and many exemplars in high frequency classes and is consistently found across corpora and languages (Evert & Baroni, 2005). The frequency spectrum indicates that the process for the present word association task is productive, meaning that if we were to sample more responses from participants the number of types would also increase.

Cue concreteness and Response frequency analyses

We computed the proportion of participants who have given a particular word in response to a particular cue to obtain the associative strength. These proportions were then multiplied by 100 to get percentage of association strength.

General statistics were then computed for the responses following a similar process as for cues. Table 2 shows the aggregated statistics for all the responses.

Table 2. Mean frequency and reaction times for response words

| | <i>M</i> | <i>SD</i> |
|--|----------|-----------|
| Movie Subtitles Frequency ^a | 29.85 | 349.18 |
| Book Frequency ^a | 33.38 | 450.23 |
| Blog Frequency ^b | 57.59 | 746.03 |
| Twitter Frequency ^b | 55.60 | 646.61 |
| Newspaper Frequency ^b | 55.26 | 797.73 |
| Reaction Times (FLP) ^c | 690.32 | 79.01 |
| Reaction Times (Megalex) ^d | 574.11 | 53.04 |

Note. a: New et al. (2004); b: Gimenes and New (2016); c: FLP: French Lexicon Project (Ferrand et al., 2010); d: Ferrand et al. (2018).

Additional analyses compared the frequency for responses elicited by abstract and concrete cue words. They revealed an effect of movie subtitle frequencies with more frequent response words corresponding to abstract cue words ($M = 87.44$; $SD = 462.69$) than more concrete cue words ($M = 72.40$; $SD = 418.31$) – $t(37146) = 3.36$; $p < 0.001$. The focus on movie subtitle frequencies was due mainly to the fact that they are commonly used, particularly when it comes to norming stimuli in language experiments².

Cue-response organisation according to concreteness

An additional study was designed to establish whether cue words prompted responses of similar concreteness levels. Using the same protocol as in Lakhzoum et al. (2021), we conducted an additional concreteness study where the responses from the 150 most strongly associated pairs were submitted to an independent group of 50 French native speakers ($M_{age} = 26.63$; $SD = 7.64$). The collected concreteness levels for both cues and responses are presented in Table A1. Results showed a strong partial correlation between cue and response concreteness levels ($r = .73$) – $t(147) = 13.00$; $p < 0.001$ controlling for number of letters,

² We used the subtitles frequency as a referent because it is the most commonly used and a better predictor compared to book frequencies for instance. This difference was not found when using book frequency.

word frequency, orthographic neighbours and Orthographic Levenshtein Distance (OLD20). An additional independent t-test showed no statistical difference between cue and response concreteness ($t(292) = 0.64$; $p = 0.52$, *ns*). Taken together, these results show that concrete cue words tend to elicit concrete responses and abstract cue words tend to elicit abstract responses. We computed the proportion of participants who have given a particular word in response to a particular cue to obtain the associative strength. These proportions were then multiplied by 100 to get percentage of association strength. Aggregating across all responses, cue-response pairs elicited by concrete cues were more strongly associated ($M = 31.0\%$; $SD = 56.0\%$) than pairs elicited by abstract cues ($M = 24.5\%$; $SD = 39.7\%$) – $t(35861) = 13.00$; $p < 0.001$.

DISCUSSION

The present study aimed to produce French word-association norms with a deliberate effort to include cues of varying levels of concreteness. The task produced a sizeable number of responses that allowed for an exploration of the dataset's distributional characteristics. Frequency spectrum analyses revealed that the dataset displayed the same pattern found across language and corpora (Evert & Baroni, 2005). This reflects the productivity of the human language and the fact that new word types are continuously produced, be it in the context of a word-association task or as a broader linguistic phenomenon.

We reported the previously collected concreteness variable for the cue words, which allowed for a comparison of response statistics between words prompted by abstract and concrete cues. The presence of the concreteness variable in this dataset, even if it is reserved for the cue words, is important especially when considering the previously reported

discrepancies in the literature as regards the role of concreteness in the organisation of the mental lexicon (see Crutch & Warrington, 2010 for example). The follow-up study we conducted to collect response concreteness revealed a high partial correlation between cue and response concreteness and no significant difference in concreteness levels between the two words of a pair. These results suggest an organisation of the mental lexicon according to a concreteness gradient. They also suggest a link between concreteness and frequency. For example, the results showed that, in comparison with concrete cues, abstract cues prompted responses that are more frequently distributed across movie subtitles. At first, this result might seem counterintuitive insofar as concrete concepts appear to be more commonly used in language. In fact, the vast majority of meaningful words used in everyday language tend to be more abstract, as illustrated by frequent abstract verbs such as *think*, *feel*, and *want* to express volition and states of mind. According to the British National Corpus (Leech, Garside & Bryant, 1994), 72 % of the noun or verb tokens in British English are more abstract than the noun *war*. Nevertheless, given that previous language norms and studies have relied heavily on concrete concepts (e.g., Barrón-Martínez & Arias-Trejo, 2014; Bonin, Méot, Ferrand & Bugańska, 2013; Ferrand & Alario, 1998), it is not surprising these results seem counterintuitive when in fact they provide more of an incentive to further the study of abstract concepts.

In a follow-up analysis we found a moderate positive correlation in perceptual strength between cues and responses ($r = .41$; $t(6466) = 35.70$; $p < 0.001$). A stronger correlation was found when considering only pairs that are strongly associated (association strength $\geq 10\%$; $r = .67$) – $t(359) = 17.00$; $p < 0.001$). This result is similar to the one obtained with the concreteness variable, suggesting modality-specific variables can also capture patterns of organisation of the mental lexicon. This result is important in light of the seeming chiasm between the distributional and grounded accounts of concept representation. This suggests the

organisation of the mental lexicon follows statistical patterns that can be captured by distributional variables, such as semantic similarity, frequency and relatedness, and by variables that are closer to the grounded end of the spectrum, such as concreteness, perceptual strength and sensorimotor norms (Lynott et al., 2020). Our analyses in respect to the latter were constrained, however, by the limited number of items between our cue-response pairs and the items present in Chedid and colleagues' (2019) database. This is a further indication of the need to develop larger normative datasets in French for modality-specific variables and with an emphasis on abstract concepts.

The aim of this dataset was to fill a gap in the French literature as regards word association norms by introducing 1 100 cues of varying levels of concreteness that elicited more than 89k responses from over 1 000 participants. To the best of our knowledge, this is the first dataset since Bonin et al. (2013), nearly a decade ago, and also the most recent since Ferrand (2001) to focus on abstract concepts. We anticipate that it will be very useful for researchers working on language- and memory-related tasks, especially those wishing to control for association when working on semantic similarity or on the impact of concreteness in the organisation of the mental lexicon.

Data Availability

The dataset for the present study is available in excel format on the OSF website (<https://osf.io/dhuqs/>). The main database is organised according to the following variables: Cue Words in French, Cue Part of Speech, Cue Word mean concreteness, Response, Frequency of the Response, total number of responses and decreasing level of Association Strength in percent.

Open Practices Statement

In line with an open data policy, all data discussed in this article are freely available on our web site on the Open Science Framework web site (<https://osf.io/dhuqs/>).

REFERENCES

- Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*, 359–370. <https://doi.org/10.1111/tops.12096>
- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498. <https://doi.org/10.1037/a0016261>
- Barrón-Martínez, J. B., & Arias-Trejo, N. (2014). Word association norms in Mexican Spanish. *The Spanish Journal of Psychology*, *17*, 1-13. <https://doi.org/10.1017/sjp.2014.91>.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660. <https://doi.org/10.1017/S0140525X99002149>
- Barsalou, L. W. (2020). Challenges and Opportunities for Grounding Cognition. *Journal of Cognition*, *3*, 31. <http://doi.org/10.5334/joc.116>
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 129–163). Cambridge University Press. <https://doi.org/10.1017/CBO9780511499968.007>
- Bonin, P., Méot, A., & Bugaiska, A. (2018). Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behavior Research Methods*, *50*, 2366–2387. <https://doi.org/10.3758/s13428-018-1014-y>
- Bonin, P., Méot, A., Ferrand, L., & Bugaiska, A. (2013). Normes d’associations verbales pour 520 mots concrets et étude de leurs relations avec d’autres variables psycholinguistiques [Verbal association norms for 520 concrete words and relationships with other psycholinguistic variables]. *L’Année Psychologique*, *113*, 63–92. <https://doi.org/10.3917/anpsy.131.0063>
- Borghini, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L.

- (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*, 263–292. <https://doi.org/10.1037/bul0000089>
- Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47. <https://doi.org/10.1613/jair.4135>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Bugajska, A., Morson, S., Moulin, C. J. A., & Souchay, C. (2011). Metamemory, recollection and familiarity in Alzheimer’s disease. *Revue Neurologique*, *167*, 3–13. <https://doi.org/10.1016/j.neurol.2010.03.001>
- Chedid, G., Brambati, S. M., Bedetti, C., Rey, A. E., Wilson, M. A., & Vallet, G. T. (2019). Visual and auditory perceptual strength norms for 3,596 French nouns and their relationship with other psycholinguistic variables. *Behavior Research Methods*, *51*, 2094–2105. <https://doi.org/10.3758/s13428-019-01254-w>
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, *33*, 497–505. <https://doi.org/10.1080/14640748108400805>
- 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*, 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*, *64*, 1388–1408. <https://doi.org/10.1080/17470218.2010.543285>
- Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*, 615–627. <https://doi.org/10.1093/brain/awh349>
- 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*, 46–71. <https://doi.org/10.1080/02643294.2010.491359>
- Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract-concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, *42*, 1042–1048. <https://doi.org/10.3758/BRM.42.4.1042>
- De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). Measuring the associative structure of English: The “Small World of Words” norms for word

- association. *Behavior Research Methods*, *51*, 987–1006. <https://doi.org/10.3758/s13428-018-1115-7>
- De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*, 480–498. <https://doi.org/10.3758/s13428-012-0260-7>
- De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*, 213–231. <https://doi.org/10.3758/BRM.40.1.213>
- De Deyne, S., Verheyen, S., & Storms, G. (2016). Structure and organization of the mental lexicon: A network approach derived from syntactic dependency relations and word associations. *Understanding Complex Systems*, *99*, 47–79. https://doi.org/10.1007/978-3-662-47238-5_3
- Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*, 284–292. <https://doi.org/10.1016/j.cognition.2008.11.012>
- Duscherer, K., & Mounoud, P. (2006). Normes d'associations verbales pour 151 verbes d'action [Word association norms for 151 action verbs]. *L'Année Psychologique*, *106*, 397–413. <https://doi.org/10.4074/S0003503306003046>
- Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*, 526–534. <https://doi.org/10.1016/j.jml.2005.01.009>
- Evert, S., & Baroni, M. (2005). Testing the extrapolation quality of word frequency models. In *Proceedings of Corpus Linguistics*. <http://www.corpus.bham.ac.uk/PCLC>
- Fernández, Á., Díez, E., Alonso, M. Á., & Beato, M. S. (2004). Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *36*, 577–583. <http://dx.doi.org/10.3758/BF03195604> https://iblues-inico.usal.es/iblues/nalc_home.php
- Ferrand, L. (2001). Normes d'associations verbales pour 260 mots « abstraits » [Word association norms for 260 'abstract' words]. *L'Année Psychologique*, *101*, 683–721. <https://doi.org/10.3406/psy.2001.29575>
- Ferrand, L., & Alario, F. X. (1998). Normes d'association verbales pour 366 noms d'objets concrets [Word association norms for 366 names of objects]. *L'Année Psychologique*, *98*, 659–709. <https://doi.org/10.3406/psy.1998.28564>
- Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., Dufau, S., Mathôt, S., & Grainger, J. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, *50*, 1285–1307. <https://doi.org/10.3758/s13428-017-0943-1>

- Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. In P. Bonin (Ed.), *The Mental Lexicon: Some Words to Talk about Words* (pp. 25–43). New-York: Nova Science Publisher.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudo words. *Behavior Research Methods*, *42*, 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547. <https://doi.org/10.1006/jmla.2000.2728>
- Firth, J. R. (1957). Applications of general linguistics. *Transactions of the Philological Society*, *56*, 1–14. <https://doi.org/10.1111/j.1467-968X.1957.tb00568.x>
- Geng, J., & Schnur, T. T. (2015). The representation of concrete and abstract concepts: Categorical versus associative relationships. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*, 22–41. <https://doi.org/10.1037/a0037430>
- Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior Research Methods*, *48*, 963–972. <https://doi.org/10.3758/s13428-015-0621-0>
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–19. <https://doi.org/10.1017/s0140525x97000010>
- Günther, F., Marelli, M., Tureski, S., & Petilli, M. A. (2022). ViSpa (Vision Spaces): A computer-vision-based representation system for individual images and concept prototypes, with large-scale evaluation. *Psychological Review*. Advance online publication. <https://psycnet.apa.org/doi/10.1037/rev0000392>
- Hamilton, A. C., & Coslett, H. B. (2008). Refractory access disorders and the organization of concrete and abstract semantics: Do they differ? *Neurocase*, *14*, 131–140. <https://doi.org/10.1080/13554790802032218>
- Harris, Z. (1954). Distributional hypothesis. *Word World*, *10*, 146–162.
- Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*, 665–695. https://doi.org/10.1162/COLI_a_00237
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*, 785–813. <https://doi.org/10.3758/BF03196544>
- Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, *4*, 103–120. <https://doi.org/10.1111/j.1756-8765.2011.01176.x>

- Joyce, T. (2005). Constructing a large-scale database of Japanese word associations. *Glottometrics, 10*, 82–99.
- Jung, J., Li, N., & Akama, H. (2010). Network analysis of Korean word associations. *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics*, 27–35. <http://www.aclweb.org/anthology/W/W10/W10-0604>
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 48*, 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>
- Kircher, T., Whitney, C., Krings, T., Huber, W., & Weis, S. (2008). Hippocampal dysfunction during free word association in male patients with schizophrenia. *Schizophrenia Research, 101*, 242–255. <https://doi.org/10.1016/j.schres.2008.02.003>
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1972). The Edinburgh Associative Thesaurus. *The Edinburgh Associative Thesaurus*. <http://vlado.fmf.uni-lj.si/pub/networks/data/dic/eat/Eat.htm>
- 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*, 14–34. <https://doi.org/10.1037/a0021446>
- Kwong, O. Y. (2013). Exploring the Chinese mental lexicon with word association norms. *27th Pacific Asia Conference on Language, Information, and Computation, PACLIC 27*, 153–162.
- Lakhzoum, D., Izaute, M., & Ferrand, L. (2021). Semantic similarity and associated abstractness norms for 630 French word pairs. *Behavior Research Methods, 53*, 1166–1178. <https://doi.org/10.3758/s13428-020-01488-z>
- 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*, 211–240. <https://psycnet.apa.org/doi/10.1037/0033-295X.104.2.211>
- Leech, G., Garside, R., & Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. *Proceedings of the 15th International Conference on Computational Linguistics*, 622–628.
- Lenci, A., Lebani, G. E., & Passaro, L. C. (2018). The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science, 10*, 550–572. <https://doi.org/10.1111/tops.12335>
- Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition, 114*, 96–104. <https://doi.org/10.1016/j.cognition.2009.09.002>
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*, 203–208. <https://doi.org/10.3758/BF03204766>

- Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*, 1271-1291. <https://doi.org/10.3758/s13428-019-01316-z>
- McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1173–1190. <https://doi.org/10.1037/0278-7393.18.6.1173>
- McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*, 558–572. <https://doi.org/10.1037/0278-7393.24.3.558>
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*, 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. ArXiv Preprint ArXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, 3111-3119.
- Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations, *Corpus Linguistics and Linguistic Theory*, *5*, 175–200. <https://doi.org/10.1515/CLLT.2009.008>
- Moss, H., Older, L., & Older, L. (1996). *Birkbeck word association norms*. Psychology Press.
- 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, and Computers*, *36*, 402-407. <https://doi.org/10.3758/BF03195588>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*, 516-524. <https://doi.org/10.3758/BF03195598>
- Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *28*, 593–599. <https://doi.org/10.1177/0963721419866441>
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25. <https://doi.org/10.1037/h0025327>
- Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, *66*, 180-194. <https://doi.org/10.1007/s00426-002->

0086-5

- Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In Cognitive Science Society (Ed.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.
- Prior, A., & Bentin, S. (2008). Word associations are formed incidentally during sentential semantic integration. *Acta Psychologica*, *127*, 57–71.
<https://doi.org/10.1016/j.actpsy.2007.01.002>
- Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*, 458-470.
<https://doi.org/10.1016/j.tics.2013.06.004>
- Qualtrics (2020). Qualtrics.com. <http://www.qualtrics.com/>
- Schloss, B., & Li, P. (2016). Disentangling narrow and coarse semantic networks in the brain: The role of computational models of word meaning. *Behavior Research Methods*, *49*, 1582–1596. <https://doi.org/10.3758/s13428-016-0807-0>
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*, 499–520. [https://doi.org/10.1016/0749-596X\(88\)90022-8](https://doi.org/10.1016/0749-596X(88)90022-8)
- Sperber, A. D., Devellis, R. F., & Boehlecke, B. (1994). Cross-cultural translation. *Journal of Cross-Cultural Psychology*, *25*, 501–524. <https://doi.org/10.1177/0022022194254006>
- Taconnat, L., Raz, N., Toczé, C., Bouazzaoui, B., Sauzéon, H., Fay, S., & Isingrini, M. (2009). Ageing and organisation strategies in free recall: The role of cognitive flexibility. *European Journal of Cognitive Psychology*, *21*, 347–365.
<https://doi.org/10.1080/09541440802296413>
- Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language*, *38*, 440–458. <https://psycnet.apa.org/doi/10.1006/jmla.1997.2559>
- Villani, C., Lugli, L., Liuzza, M. T., Nicoletti, R., & Borghi, A. M. (2021). Sensorimotor and interoceptive dimensions in concrete and abstract concepts. *Journal of Memory and Language*, *116*, 104-173. <https://doi.org/10.1016/j.jml.2020.104173>
- West, J. T., & Mulligan, N. W. (2019). Prospective metamemory, like retrospective metamemory, exhibits under confidence with practice. *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*, 22-24.
<https://doi.org/10.1037/xlm0000708>
- Wingfield, C., & Connell, L. (2022a). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, 1-51.
<https://doi.org/10.1080/23273798.2022.2069278>

- Wingfield, C., & Connell, L. (2022b). Sensorimotor distance: A grounded measure of semantic similarity for 800 million concept pairs. *Behavior Research Methods*, 1-17. <http://dx.doi.org/10.3758/s13428-022-01965-7>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin and Review*, 15, 971–979. <https://doi.org/10.3758/PBR.15.5.971>
- Zwaan, R. A. (2004). The Immersed Experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation - Advances in Research and Theory*, 44, 35–62. [https://doi.org/10.1016/S0079-7421\(03\)44002-4](https://doi.org/10.1016/S0079-7421(03)44002-4)

APPENDIX

Table A1. First 150 most associated pairs with cue and response mean concreteness

| cue | translation | response | translation | Assoc_ strength (%) | cue_concreteness | resp_concreteness |
|--------------|---------------|----------|-------------|------------------------|------------------|-------------------|
| crainte | fright | peur | fear | 85.71 | 2.95 | 2.37 |
| omelette | omelet | œuf | egg | 82.05 | 6.67 | 6.61 |
| chorégraphie | choregraphy | danse | dance | 79.49 | 5.12 | 4.71 |
| poivre | pepper | sel | salt | 78.57 | 6.55 | 6.10 |
| croquis | sketch | dessin | drawing | 74.39 | 5.46 | 5.57 |
| ruche | hive | abeille | bee | 72.73 | 6.43 | 6.67 |
| brasier | inferno | feu | fire | 72.37 | 5.75 | 5.76 |
| question | question | réponse | answer | 72.00 | 3.93 | 2.94 |
| homme | man | femme | woman | 71.88 | 6.26 | 5.78 |
| selle | saddle | cheval | horse | 71.43 | 6.46 | 6.63 |
| luge | sledge | neige | snow | 70.42 | 6.58 | 6.29 |
| primate | primate | singe | monkey | 70.27 | 5.36 | 6.43 |
| soucoupe | saucer | volant | flying | 69.77 | 5.90 | 5.61 |
| égaré | misplaced | perdre | loose | 69.57 | 3.49 | 2.86 |
| prévision | forecast | météo | weather | 69.57 | 2.74 | 3.88 |
| avalanche | avalance | neige | snow | 69.14 | 6.17 | 6.29 |
| aéroport | airport | avion | plane | 67.95 | 6.24 | 6.63 |
| nid | nest | oiseau | bird | 67.86 | 6.31 | 6.69 |
| effroi | dread | peur | fear | 66.67 | 3.16 | 2.37 |
| robinet | faucet | eau | water | 65.52 | 6.61 | 6.22 |
| tendance | tendency | mode | fashion | 64.84 | 2.57 | 2.82 |
| rouge-gorge | robin | oiseau | bird | 64.71 | 6.58 | 6.69 |
| cacao | cocoa | chocolat | chocolate | 64.37 | 6.23 | 6.59 |
| ancre | anchor | bateau | boat | 62.35 | 6.14 | 6.63 |
| chlore | chlorine | piscine | pool | 62.12 | 5.41 | 6.49 |
| bibliothèque | library | livre | book | 61.97 | 6.48 | 6.63 |
| baril | barrel | pétrole | oil | 61.33 | 6.34 | 5.88 |
| périr | perish | mourir | die | 61.29 | 4.11 | 3.51 |
| riche | wealthy | pauvre | poor | 61.29 | 3.48 | 2.94 |
| flamme | flame | feu | fire | 61.11 | 5.94 | 5.76 |
| réponse | answer | question | question | 60.53 | 3.92 | 3.00 |
| cliché | stereotype | photo | picture | 60.44 | 3.71 | 5.73 |
| bravoure | bravery | courage | courage | 60.26 | 2.97 | 2.69 |
| cercle | circle | rond | round | 60.24 | 5.21 | 4.82 |
| ampoule | bulb | lumière | light | 59.55 | 6.78 | 4.76 |
| gare | train station | train | train | 59.46 | 6.19 | 6.55 |
| frère | brother | sœur | sister | 59.21 | 5.94 | 5.67 |
| roi | king | reine | queen | 59.15 | 5.59 | 5.31 |

| | | | | | | |
|-------------|-------------|------------|------------|-------|------|------|
| oncle | uncle | tante | aunt | 58.82 | 6.46 | 5.75 |
| pacifique | peaceful | océan | ocean | 58.75 | 3.64 | 5.82 |
| bien | good | mal | bad | 58.11 | 2.60 | 2.22 |
| décret | decree | loi | law | 57.97 | 4.26 | 3.04 |
| tablier | apron | cuisine | kitchen | 57.69 | 6.53 | 5.73 |
| tonneau | cask | vin | wine | 57.69 | 6.63 | 6.35 |
| vie | life | mort | death | 57.69 | 3.74 | 3.45 |
| femme | woman | homme | man | 57.33 | 5.86 | 5.57 |
| grand | large | petit | small | 57.14 | 4.59 | 3.88 |
| soulier | shoe | chaussure | shoe | 56.99 | 6.69 | 6.45 |
| réfléchir | reflect | penser | think | 56.63 | 3.62 | 2.82 |
| moineau | sparrow | oiseau | bird | 56.47 | 6.48 | 6.69 |
| célébration | celebration | fête | party | 55.95 | 3.97 | 4.06 |
| égalité | equality | fraternité | fraternity | 55.95 | 3.05 | 2.49 |
| étincelle | spark | feu | fire | 55.26 | 5.58 | 5.76 |
| après | after | avant | before | 55.13 | 3.77 | 2.45 |
| terrier | burrow | lapin | rabbit | 54.67 | 6.43 | 6.65 |
| monarchie | monarchy | roi | king | 53.73 | 4.04 | 5.43 |
| berceau | crib | bébé | baby | 53.41 | 6.60 | 6.20 |
| anatomie | anatomy | corps | body | 53.09 | 5.82 | 5.96 |
| sapin | fir | noël | christmas | 52.63 | 6.76 | 3.88 |
| trébucher | stumble | tomber | fall | 52.63 | 5.26 | 4.12 |
| capture | capture | écran | screen | 52.44 | 3.75 | 6.41 |
| rusé | cunning | renard | fox | 52.05 | 3.27 | 6.73 |
| lait | milk | vache | cow | 51.81 | 6.48 | 6.67 |
| astronomie | astronomy | étoile | stars | 51.76 | 4.59 | 5.59 |
| émissaire | emissary | bouc | goat | 51.47 | 4.33 | 6.29 |
| abdomen | abdomen | ventre | belly | 51.32 | 6.53 | 6.14 |
| privé | private | public | public | 51.28 | 3.31 | 3.84 |
| bulle | bubble | savon | soap | 51.16 | 5.71 | 6.43 |
| flore | flora | faune | fauna | 50.77 | 5.24 | 4.47 |
| existence | existence | vie | life | 50.55 | 3.52 | 3.14 |
| palpitation | palpitation | coeur | fever | 50.53 | 4.87 | 5.47 |
| album | album | photo | photo | 50.00 | 6.23 | 5.73 |
| tentative | attempt | essai | test | 50.00 | 3.37 | 3.35 |
| navire | ship | bateau | boat | 49.44 | 6.51 | 6.63 |
| problème | problem | solution | solution | 49.41 | 3.47 | 3.10 |
| tasse | cup | café | coffee | 49.37 | 6.68 | 6.43 |
| source | source | eau | water | 49.30 | 4.20 | 6.22 |
| liquide | liquide | eau | water | 49.21 | 5.34 | 6.22 |
| torchon | napkin | serviette | towel | 48.84 | 6.41 | 6.47 |
| panique | panic | peur | fear | 48.78 | 3.49 | 2.37 |
| brise | breeze | vent | wind | 48.65 | 4.51 | 4.92 |
| rectangle | rectangle | carré | square | 48.61 | 6.56 | 3.55 |

| | | | | | | |
|-------------|---------------|-----------|-------------|-------|------|------|
| pauvre | poor | riche | wealthy | 48.53 | 3.83 | 3.25 |
| Pâque | Easter | œuf | egg | 48.44 | 3.97 | 6.61 |
| bizarre | weird | étrange | strange | 48.24 | 2.50 | 2.27 |
| épave | wreck | bateau | boat | 47.89 | 5.80 | 6.63 |
| dévoré | devour | manger | eat | 47.76 | 4.20 | 4.82 |
| démence | dementia | folie | madness | 47.56 | 2.87 | 2.10 |
| étrange | strange | bizarre | weird | 47.52 | 2.49 | 2.02 |
| bouée | buoy | sauvetage | rescue | 47.50 | 6.51 | 3.82 |
| forêt | forest | arbre | tree | 47.44 | 6.00 | 6.67 |
| chouette | owl | hibou | owl | 47.37 | 5.97 | 6.49 |
| aisé | convenient | riche | wealthy | 47.13 | 3.26 | 3.25 |
| solution | solution | problème | problem | 46.99 | 3.91 | 2.80 |
| coïncidence | coincidence | hasard | chance | 46.91 | 2.77 | 2.27 |
| hibou | owl | chouette | owl | 46.48 | 6.78 | 5.53 |
| pinceau | brush | peinture | paint | 46.39 | 6.56 | 5.92 |
| arrivée | arrival | départ | departure | 46.38 | 4.07 | 3.51 |
| soi | self | moi | me | 46.34 | 2.84 | 4.12 |
| avoir | to have | être | to be | 46.25 | 3.49 | 2.08 |
| entrée | entrance | sortie | exit | 45.65 | 4.56 | 3.67 |
| geler | congeal | froid | cold | 45.65 | 5.05 | 3.80 |
| timbre | stamp | poste | post office | 45.59 | 6.35 | 4.80 |
| planète | planet | terre | earth | 45.56 | 5.68 | 5.98 |
| célébrité | fame | star | star | 45.35 | 3.33 | 4.06 |
| bon | good | mauvais | bad | 45.26 | 3.13 | 2.51 |
| monter | go up | descendre | come down | 45.24 | 4.50 | 3.82 |
| soir | evening | nuit | night | 45.00 | 4.84 | 2.63 |
| auteur | author | livre | book | 44.93 | 5.54 | 6.63 |
| poignée | handle | porte | door | 44.59 | 5.95 | 6.47 |
| vache | cow | lait | milk | 44.58 | 6.89 | 6.45 |
| budget | budget | argent | money | 44.44 | 5.03 | 5.24 |
| finance | finance | argent | money | 44.30 | 3.45 | 5.24 |
| cage | cage | oiseau | bird | 44.29 | 6.31 | 6.69 |
| collège | middle school | lycée | high school | 44.29 | 6.17 | 5.71 |
| divertir | entertain | amuser | amuse | 43.94 | 3.35 | 2.88 |
| compte | account | banque | bank | 43.90 | 4.44 | 5.47 |
| individu | individual | personne | person | 43.90 | 5.10 | 4.71 |
| mauvais | bad | bon | good | 43.84 | 3.26 | 2.71 |
| mélancolie | melancoly | tristesse | sadness | 43.84 | 2.69 | 2.61 |
| béton | concrete | armer | reinforced | 43.68 | 6.52 | 3.92 |
| propre | clean | sale | dirty | 43.62 | 3.98 | 3.73 |
| cartable | schoolbag | école | school | 43.59 | 6.78 | 5.96 |
| pyramide | pyramid | Egypte | Egypt | 43.53 | 6.31 | 5.31 |
| avenir | future | futur | future | 43.48 | 3.04 | 2.27 |
| plage | beach | sable | sand | 43.43 | 6.09 | 6.49 |

| | | | | | | |
|-----------|----------|----------|-----------|-------|------|------|
| velouté | chowder | soupe | soup | 43.43 | 5.35 | 6.33 |
| faiblesse | weakness | force | strength | 43.42 | 2.72 | 3.04 |
| neveu | nephew | nièce | niece | 43.18 | 6.20 | 5.67 |
| pays | country | France | France | 42.86 | 5.66 | 4.94 |
| maximum | maximum | minimum | minimum | 42.71 | 3.04 | 2.96 |
| mal | bad | bien | good | 42.47 | 3.14 | 2.53 |
| ricaner | giggle | rire | laugh | 42.47 | 4.51 | 3.90 |
| berger | shepherd | mouton | sheep | 42.45 | 6.10 | 6.65 |
| aérien | aerial | avion | plane | 42.19 | 4.21 | 6.63 |
| action | action | réaction | reaction | 42.17 | 4.09 | 2.86 |
| poignard | dagger | couteau | knife | 42.17 | 6.39 | 6.61 |
| lit | bed | dormir | sleep | 41.98 | 6.84 | 4.37 |
| adulte | adult | enfant | child | 41.94 | 4.40 | 5.86 |
| émeraude | emerald | vert | green | 41.94 | 6.21 | 4.37 |
| gravir | climb | monter | ascend | 41.94 | 4.94 | 3.90 |
| courgette | zucchini | légume | vegetable | 41.46 | 6.76 | 6.12 |
| abeille | bee | miel | honey | 41.38 | 6.64 | 6.33 |
| printemps | spring | été | summer | 41.38 | 4.38 | 3.80 |
| orne | adorn | décorer | decorate | 41.05 | 4.61 | 3.82 |
| horreur | horror | film | movie | 41.03 | 3.00 | 5.69 |
| célébrer | showcase | fêter | celebrate | 40.82 | 3.89 | 3.75 |
| part | piece | gâteau | cake | 40.74 | 3.82 | 6.31 |
| colonie | colony | vacance | vacation | 40.63 | 4.30 | 3.73 |