Originally published in: SAGE Open vol. 14 (2024) no. 1, 11 pp.
DOI: *https://doi.org/10.1177/21582440231219101*

Original Research

SAGE Open
January-March 2024: 1–11
© The Author(s) 2024
DOI: 10.1177/21582440231219101
journals.sagepub.com/home/sgo
Sage

What Lexical Factors Drive Look-Ups in the English Wiktionary?

Robert Lew[1] and Sascha Wolfer[2]

Abstract

This study aims to establish what lexical factors make it more likely for dictionary users to consult specific articles in a dictionary using the English Wiktionary log files, which include records of user visits over the course of 6 years. Recent findings suggest that lexical frequency is a significant factor predicting look-up behavior, with the more frequent words being more likely to be consulted. Three further lexical factors are brought into focus: (1) age of acquisition; (2) lexical prevalence; and (3) degree of polysemy operationalized as the number of dictionary senses. Age of acquisition and lexical prevalence data were obtained from recent published studies and linked to the list of visited Wiktionary lemmas, whereas polysemy status was derived from Wiktionary entries themselves. Regression modeling confirms the significance of corpus frequency in explaining user interest in looking up words in the dictionary. However, the remaining three factors also make a contribution whose nature is discussed and interpreted. Knowing what makes dictionary users look up words is both theoretically interesting and practically useful to lexicographers, telling them which lexical items should be prioritized in lexicographic work.

Plain Language Summary

What makes people look up words in the English Wiktionary?

This study aims to establish what factors make it more likely for dictionary users to consult specific articles in a dictionary using the English Wiktionary log files, which include records of user visits over the course of six years. Recent findings suggest that word frequency is a significant factor predicting look-up behaviour, with the more frequent words being more likely to be consulted. Three further factors are brought into focus: (1) age of acquisition, which is the age at which a word is learned; (2) lexical prevalence, which is how many people know the word; and (3) degree of polysemy calculated as the number of dictionary senses. Age of acquisition and lexical prevalence data were obtained from recent published studies and linked to the list of visited Wiktionary lemmas, whereas polysemy status was derived from Wiktionary entries themselves. Our study confirms the significance of word frequency in explaining user interest in looking up words in the dictionary. However, the remaining three factors also make a contribution whose nature is discussed and interpreted. Knowing what makes dictionary users look up words is both theoretically interesting and practically useful to lexicographers, telling them which words should be prioritized in lexicographic work.

Keywords

lexicography, language studies, humanities, log files, Wiktionary, online dictionary, corpus frequency, dictionary

Introduction

The Role of Dictionaries in Today's World

Dictionaries have been part and parcel of literate societies for many centuries. The most prominent role of dictionaries in society has been to assist in communication—be it in one language or across different languages—to aid in understanding, creating, and translating texts. Thousands of languages are spoken around the world, and communication problems arise whenever a native speaker of one

[1]Adam Mickiewicz University, Poznań, Poland
[2]Leibniz-Institut für Deutsche Sprache, Mannheim, Germany

Corresponding Author:
Robert Lew, Faculty of English, Adam Mickiewicz University, Wieniawskiego 1, Poznań 61-712, Poland.
Email: rlew@amu.edu.pl

Data Availability Statement included at the end of the article

language comes into contact with a speaker of another language. In today's *global village*—marked by the ubiquity of long-distance travel, increased human mobility, and modern communication technology (the Internet and mobile telephony)—frequent contacts between speakers of different languages have become a major part of our daily experience. At the same time, English has established itself as a lingua franca of international communication: the language of choice in communication between people speaking different languages natively. This marked tendency gives lexicography of English a particular significance, as dictionaries with English are used intensively and extensively by huge numbers of people worldwide. For the English Wiktionary—the dictionary we are using as our primary data source here—the relevance of English language resources around the world is reflected in the page impression statistics for different countries[1]: the majority of page impressions (52.7%) originate from countries other than the five countries with the most native speakers of English[2] (USA, Great Britain, Canada, Nigeria, and Australia).

The Role of Corpora in Lexicography

In the not-so-distant past, lexicographers conceived and compiled dictionaries by relying on primarily two sources of data: their own introspection and past practice. Because of its reliance on introspection and largely uncritical copying from earlier dictionaries, pre-modern lexicography was anti-empirical as well as strongly conservative. Empirical evidence first came into lexicography in the 19th century with citation slips and reading programs (Atkins & Rundell, 2008, p. 50). This method was prone to human bias, as humans tend to notice the unusual, and ignore the habitual. More objectivity was only brought in with the corpus revolution pioneered by the COBUILD project (Sinclair, 1987).

Introduction of systematic corpus data was important in that it offered information, among other things, on which words are most frequent in the language, to the extent that the corpora employed were representative of the language. This allowed lexicographers—always contending with limited resources and time strictures—to focus their efforts on words that corpus analysis found to be most frequently used. In this fashion, corpora offered relatively objective data on language use, but had nothing to say about how people used dictionaries, or to what extent their interest in words aligned with corpus data. In particular, there was no telling whether words found to be frequent text-wise were also the ones that people wanted to consult most often in the dictionary. For that kind of insight, lexicographers needed real data on dictionary use.

Dictionary Log Files

The study of how people use dictionaries began in the late 20th century, and mostly consisted in recording a limited group of study subjects in the process of dictionary use, or in looking at the product of such dictionary use, such as lexical choices, sentences, or texts produced with the help of dictionaries. While this approach offered details relevant to the design of entry structure and organization, the low volume of the data as well as the frequent artificiality of the context and tasks meant that there was little information there on which words dictionary users would normally wish to look up in dictionaries under naturalistic circumstances.

New opportunities to collect such information on a larger scale came with the transition of lexicography to the digital medium (Lew & de Schryver, 2014), as some digital dictionaries, particularly those accessible online, log details of user visits. Such logs may be subsequently mined for information on which parts of the dictionary were accessed and with what frequency. Such data will sometimes be used by the publisher, though not usually shared outside due to their commercial value. However, for English there exists the popular and substantial English Wiktionary, which is a non-commercial crowd-sourced resource. For this dictionary, extensive log files are available for download, thus providing an excellent opportunity for study.

Potentially Relevant Lexical Factors

Corpus Frequency.

Corpus-based lexical frequency has been an important consideration in determining the coverage of a dictionary, beginning with the pioneering work completed as part of the COBUILD project (Sinclair, 1987). Still, quite surprisingly, the positive relationship between dictionary look-up and corpus frequency did not turn out to be apparent at all in early studies looking into this issue (De Schryver & Joffe, 2004; De Schryver et al., 2006; Verlinde & Binon, 2010), and has only been established empirically with some confidence fairly recently (De Schryver et al., 2019; Koplenig et al., 2014; Müller-Spitzer et al., 2015), although frequency-based heuristics had been tested much earlier in a contrastive setting with two dictionaries (Verlinde & Selva, 2001). Armed with more sophisticated data analysis tools unavailable to early studies, De Schryver et al. (2019) found a clear positive relationship between corpus frequency and user interest, indicating that words with higher corpus frequency tend to be more frequently looked up by users. This effect was found for both English and Swahili in an online English-Swahili dictionary.

Word frequency also turns out to be an important factor in predicting behavior in reading text (Kliegl et al., 2006), lexical decision tasks (Morrison & Ellis, 1995), and a wide range of other tasks during language processing (N. C. Ellis, 2002). Another concept that is very tightly linked to corpus frequency is orthographic familiarity (White, 2008) which is an operationalization of how often a word with a similar "shape" (same initial letters and same length) as the target word can be observed in every-day language.

However, researchers are increasingly looking—especially in psycholinguistics if not so much (yet) in lexicography (Brysbaert et al., 2018, 2019; Mandera et al., 2017, 2020)—at the possibility that there are other non-trivial aspects of *word knowledge*, beyond mere frequency of occurrence, possibly playing a role in how "interesting" a word is found by speakers (Bialystok et al., 2009; Gerhand & Barry, 1998; Goodman et al., 2008; Mandera et al., 2015). We would like to explore this avenue insofar as it is evidenced in dictionary look-up behavior. Whilst the relationship between corpus frequency and look-up behavior has received some attention, we see a clear advantage in including further variables. Additional metrics describing other properties of words (some of them closely related—but not identical—to corpus frequency) can also help us understand better the effect of corpus frequency and the relationships between predictor variables. Three of those candidate predictors that appear most promising and will be examined in this contribution are briefly introduced below.

Word Prevalence. The prevalence of a word is the extent to which it is known amongst the native-speaking population. Words which occur with relatively higher frequency in texts and discourse should be more likely to be known by a large proportion of speakers (Longobardi et al., 2015; Weizman & Snow, 2001). Conversely, it would not be reasonable to expect quite rare words to be known to a broad majority of the speakers of a language. All this does not, however, preclude words of moderate frequency being more or less widely known, perhaps due to the relative ubiquity of the concepts that some of them might convey. Another complication is distribution across texts of different type, modality, or genre: words can be frequent, but with most tokens concentrated in a limited range of texts, which would not be conducive to universal prevalence. The state-of-the-art approach to collecting word prevalence information is through crowdsourcing, employing large-scale online surveys (Brysbaert et al., 2019), asking speakers if they are familiar with specific words.

Age of Acquisition. Age of acquisition is the age at which a word is, on average, acquired by native speakers in the process of (naturalistic) L1 acquisition. One might expect that this could play a role in how words acquired earlier, possibly being more deeply entrenched in the mental lexicon, get to be looked up. Age of acquisition has been found to have important and long-lasting effects on language behavior (A. W. Ellis & Lambon Ralph, 2000; Garlock et al., 2001; Juhasz, 2005; Kuperman et al., 2012; Morrison et al., 1992; Weizman & Snow, 2001). Of special interest is a study by Navarrete et al. (2015), which suggests that words that are acquired later in life are more likely to elicit "tip-of-the-tongue" phenomena (a sensation in which a known word is momentarily inaccessible). It is not unlikely that such a state could trigger dictionary consultation. Another study by Picard et al. (2010) suggests that the "core" of the dictionary, that is, words that are strongly interconnected and are key when it comes to learning a new language, are the ones that native speakers tend to acquire at a significantly younger age.

Number of Senses (Degree of Polysemy). Words can (and often do) have more than one meaning, or sense. The concept of *word sense* is not without problems (Hanks, 2000; Kilgarriff, 1997), and there has been a long-drawn-out debate about the boundaries between polysemy and homonymy. Some lexicographers even explicitly identify as either *lumpers* or *splitters* (Van der Meer, 2004). To steer clear of the essentialist debate of whether words "have" senses, we will adopt a pragmatic approach of considering lexicographic senses, that is, the separate blocks of meaning description as given in a dictionary, in our case operationalized as the number of dictionary senses in the English Wiktionary. We have known for more than 70 years (Zipf, 1949) that the more frequent words tend to have more senses. However, the degree of polysemy may hold predictive potential above and beyond that of mere word frequency.

To wrap up this overview section, Table 1 summarizes the essential parameters of previously published log-file-based studies examining the role of lexical factors in dictionary consultation, along with analogous data for the present study.

Aim

As argued above, when people elect to use dictionaries, they make choices about which words to look up, and our present aim is to try to identify the lexical variables that affect the likelihood of those choices by using the log files of a popular crowd-sourced dictionary: the English Wiktionary. At the same time, there is some controversy as to the relationship between lexical frequency and dictionary user look-up frequency, with some studies finding no such relationship, and others reporting a positive relationship. While at this time this difference in findings appears to be in the use of more sophisticated methods of exploring this relationship, we still see the need to confirm the findings that the more frequent words are indeed

Table 1. Parameters of Previous Log-File Studies of the Role of Lexical Factors in Dictionary Acquisition Compared to the Present Study.

Reference	Lexical factors considered	Dictionaries studied	Data size (when reported)	Main findings
De Schryver and Joffe (2004)	Frequency	Sesotho sa Leboa Dictionary Project (SeDiPro)	25K Sesotho items, 28K English search terms 21K look-ups	Positive effect of frequency
De Schryver et al. (2006)	Frequency	Online Swahili-English Dictionary	6.5K Swahili items, 11.5K English search terms Half a million look-ups	Slight positive frequency effect at highest frequencies
Verlinde and Binon (2010)	Frequency	Base lexicale du français	56K look-ups	Weak positive effect of frequency
Koplenig et al. (2014)	Frequency	Digitales Wörterbuch der deutschen Sprache (DWDS) + German Wiktionary	581K DWDS look-ups + 1.6M Wiktionary look-ups	Positive effect of frequency
Müller-Spitzer et al. (2015)	Frequency + polysemy	German Wiktionary	71K entries; 82M look-ups	Positive effects of frequency and polisemy
De Schryver et al. (2019)	Frequency	Online Swahili-English Dictionary	15M Swahili look-ups + 10M English look-ups	Positive effect of frequency
Present study	Frequency + prevalence + age of acquisition + polysemy + POS (part of speech)	English Wiktionary	31K entries; 780M look-ups	Positive effects of frequency, polysemy, and age of acquisition; negative effect of prevalence; some effect of POS

looked up more often. However, even if lexical frequency is a useful predictor, it seems clear that other factors are involved. While not seeking a single lexical processing or representation model, we are interested in what drives people's decisions to look up a specific word in terms of language experience. This leads us to the following research question with four sub-questions:

- How do the following lexical factors affect dictionary users' decisions to look up specific words:
  1. corpus frequency (verify the positive relationship)
  2. word prevalence
  3. age of acquisition
  4. degree of polysemy

Methodology

Data Sources and Data Integration

Age-of-acquisition (AoA) ratings and corpus frequencies were extracted from the supplementary material made available by Kuperman et al. (2012). AoA ratings are represented as the average rating of 1,960 responders on Amazon Mechanical Turk. Kuperman et al. (2012) show that their data of 842,438 ratings "are as valid and reliable as those collected in laboratory conditions" (p. 978).

Corpus frequency is given as standardized frequency values expressed as hits per 1 million tokens, and are computed from raw frequency figures given in the SUBTLEX-US corpus (as described in Brysbaert & New, 2009).

Prevalence values were extracted from the supplementary material published as part of Brysbaert et al. (2019). The prevalence data are based on the authors' original web-based survey and covered responses from a total of 221,268 English-speaking participants living in the US and UK; the complete dataset of prevalence values comprises 61,855 data points.

For information on polysemy, we extracted the number of senses for each word directly from its dictionary entry in the English Wiktionary itself. For this, we used a custom R (R Core Team, 2022) function which accesses the edit page of each article. For example, for the entry "dictionary," the URL https://en.wiktionary.org/w/index.php?title=dictionary&action=edit is being scraped by the function.[3] The extraction script is available upon request.

We extracted part-of-speech (POS) information directly for each word from its English Wiktionary entry, just as we did for the number of senses. Here too we used a custom R function (also available upon request), but this time we used the entry page itself for extraction. Our final dataset includes two types of POS information: (1) a list of *all* parts-of-speech given in the entry; and (2) the *first* part-of-speech listed. There were a total of four entries in our dataset (*disrobement, iceskate, liquescence*, and *polloi*) for which we had to assign POS information

Table 2. Information on the Variables Used in the Present Study.

Variable name	Transformation(s)	Short name	Data type	Source/reference
Number of look-ups	log10	log.views	Continuous	R package pageviews
Age of acquisition	Standardized	std.AOA	Continuous	Kuperman et al. (2012)
Corpus frequency	Standardized, log10	std.log.Freq	Continuous	SUBTLEX-US (Brysbaert & New, 2009)
Prevalence	Standardized	std.Prev	Continuous	Brysbaert et al. (2019)
Polysemy	None	polysem	Binary	Wiktionary pages (custom R function)
Part-of-speech group	None	first.pos.cat	Categorial (4 groups)	Wiktionary pages (custom R function)

Table 3. Results for the Linear Regression Model.

Predictor	Estimate	*t*-Value	*p*-Value	VIF	ΔR^2
Standardized AoA	0.255	34.62	<.00001	1.853	.019
Standardized log frequency per one million tokens	1.020	131.17	<.00001	2.098	.272
Standardized prevalence	−0.147	−19.52	<.00001	1.832	.006
Polysemy (T/F)	0.736	57.72	<.00001	1.139	.053

Note. Predictor: name of the predictor; Estimate: beta estimate from the linear regression model; *t*-value: associated *t* value (Estimate divided by associated standard error); *p*-value: indicator of statistical significance; VIF: variance inflation factor for the predictor; ΔR^2: difference between full model R^2 and R^2 of a model without the predictor (see further explanation in the text).

manually due to the different structure of these entries. For three entries, we had to extract POS information from their spelling variants (*Capricorn, gokart*, and *plutonian*).

The criterion we attempt to predict using the above variables is the number of look-ups for each of the online entries. To collect this information, we used the R package pageviews (Keyes & Lewis, 2020). We restricted page view information to non-automatic look-ups. Separate look-up counts are available for desktop access, mobile access via a web browser, and mobile access via the Wiktionary app. The time span that we collected daily look-up data for is 01-01-2016 to 31-10-2021. Aggregated figures are available for monthly, yearly, and overall look-ups for each dictionary entry.

To integrate all data described above in one single dataset, we first identified all intersecting lexical items from the AoA, frequency, and prevalence lists. We then checked which of these items had a corresponding entry in the English Wiktionary. For each of these entries, we extracted sense, POS, and look-up data. All in all, our final dataset contains approximately 780 million look-ups distributed over 30,750 entries. Table 2 gives an overview of all the variables in our dataset relevant for the present study.

Data Analysis

Data Transformation. The distribution of look-up data (as measured by the number of views for each article) is heavily skewed toward low values. After log-transforming, the variable approaches normal distribution, and so we adopted the transformed variable (log views) as the criterion variable in our statistical model. Likewise, we log-transformed the standardized corpus frequency predictor variable to normalize the distribution of the residuals of the linear model, as per a general assumption of linear regression models.

We then standardized all continuous predictors (AoA, log frequency, and prevalence) to *z*-scores by subtracting the respective mean from each value and then dividing by the respective standard deviation. This maps all the continuous predictors on the same scale with a mean value of 0 and a standard deviation of 1, making linear regression model estimates comparable.[4]

Model Specification. We predicted log views by standardized AoA, log frequency, and prevalence. Polysemy (true/false) entered the model as a categorial predictor. The corresponding R formula is:

$$\text{log.views} \sim \text{std.AOA} + \text{std.log.Freq} + \text{std.Prev} + \text{polysem}$$

The coefficient of determination for the full model is $R^2 = .5228$. Further details on this model are given in Table 3.

To spot potential problems with collinearity in the model (for example, log frequency and prevalence are correlated at $r_{\text{Pearson}} = .62$ and $r_{\text{Spearman}} = .72$), we checked the variance inflation factors (VIF) of the predictors. As can be seen in Table 3, none of the VIFs approaches or exceeds a value (5 or 10) that could indicate “a problematic amount of collinearity” (James et al., 2013, p. 101f).

We also tested whether model results are crucially influenced by outliers in the log views, our criterion variable. After excluding 24 data points (i.e., less than 0.1% of all data points) lying outside the hinges of a boxplot with $r = 1.5$ (which is a rather strict criterion), none of the effects reported above changed in a meaningful manner, that is, the overall effect pattern stayed the same.

We do not present models with any interactions here. However, we also computed an alternative model which included all possible two-way interactions. Several of these interaction terms did not reach statistical significance and were excluded. Another interaction effect was excluded because it led to inflated variance in the model (highest VIF: 9.22). After this, two two-way interaction terms remained in the alternative model (AoA:Prevalence and AoA:Polysemy). Even so, as the gain in explained variance from including these two interaction terms was close to none ($R^2_{\mathrm{int}} = .5247$, compared to $R^2 = .5228$ for the original no-interaction model), there was little justification for retaining the two additional terms in the model. For reference, we include this alternative model in the Supplemental Material.

A Preliminary Look at Part of Speech Labels.

In a separate preliminary analysis in response to a suggestion by an anonymous reviewer, we investigated whether look-ups vary by part of speech. We did not include POS as a predictor in the regression model, because the POS information is not as reliable as the other predictors: here, we only used the *first* part-of-speech label on the entry page in the English Wiktionary. In some cases, this seems rather arbitrary. For example, the first POS given at the entry “a” is “Letter,” whereas “Article” would have been a more appropriate category to represent the most salient use of the word “a.” We grouped these POS labels into the following four categories (starting with the most numerous group): (1) Nouns and proper nouns ($n = 19{,}258$); (2) Adjectives and adverbs ($n = 7{,}689$); (3) Verbs ($n = 3{,}577$); and Others[5] ($n = 226$). In this preliminary analysis, we compared the (log-transformed) views with pairwise t-tests between all groups.

Results

All predictors are highly significant in their contribution toward predicting article views in the English Wiktionary. Frequency and AoA show a positive relationship with article views. This means that words that are found more often in a corpus tend to be looked up more often, and words that are acquired later in life are also more likely to receive more views.

Words that are more prevalent in the population, that is, are known to more people, are looked up *less* often. This is indicated by the negative estimate for the standardized prevalence variable in Table 3.

To assess the relative importance of the predictors, we can refer to absolute values of the continuous predictors thanks to their prior standardization. This shows a clear hierarchy of importance: corpus frequency is by far the most important predictor in the model, followed by AoA and prevalence (recall that polysemy is not a continuous predictor). As a second measure of relative importance, we refer to the proportion of variance in the criterion that is explained by the predictors in the model (R^2). Here, we drop each variable from the model and calculate the difference between the full model (as indicated above: $R^2 = .5228$) and the model *without* the respective variable and call this measure ΔR^2. Conversely, ΔR^2 can also be interpreted as a measure of how much more variance in the number of views is explained by the model if the respective predictor is included. As expected, this shows the same hierarchy as the comparison of estimates. In addition, we can include polysemy in the hierarchy because ΔR^2 does not rely on standardized predictors.

Figure 1 visualizes all effects from the linear regression model. The importance ranking of the variables is here also apparent in the range spanned by the values of predicted views. For example, the very strong effect of frequency leads to an increase from near-zero views to nearly 2 million predicted views for entries for very frequent words. In contrast, the effect of prevalence is visually apparent, but from one end of the standardized prevalence scale to the other, predicted views only change by a factor of 2.

Figure 2 shows the distribution of the (log-transformed) look-ups of entries grouped in four POS categories. Pairwise t-tests indicate highly significant differences (all Holm-adjusted p-values $< .0001$) between all groups except between (proper) nouns and verbs ($p = .059$). It is quite obvious that especially the “Others” category stands out from the rest. We propose two alternative reasons for this. The “Others” group contains high-frequency function words that are also looked up very often (e.g., “a,” “I,” “what,” “the,” “for,” “in,” “more”). Alternatively, the size of the group could lie at the heart of the difference: while the other groups contain several thousand entries each, only 220 words fall into the “Others” category. That is because this category captures non-productive, closed syntactic classes of the vocabulary (unlike nouns, verbs, or adjectives). But this also means fewer entries that could drag the distribution of look-ups down. To illustrate this point: the lower end of the rightmost violin (“Others”) in Figure 2 is at 2,469 look-ups for the entry “huzza” (POS tag “Interjection”).

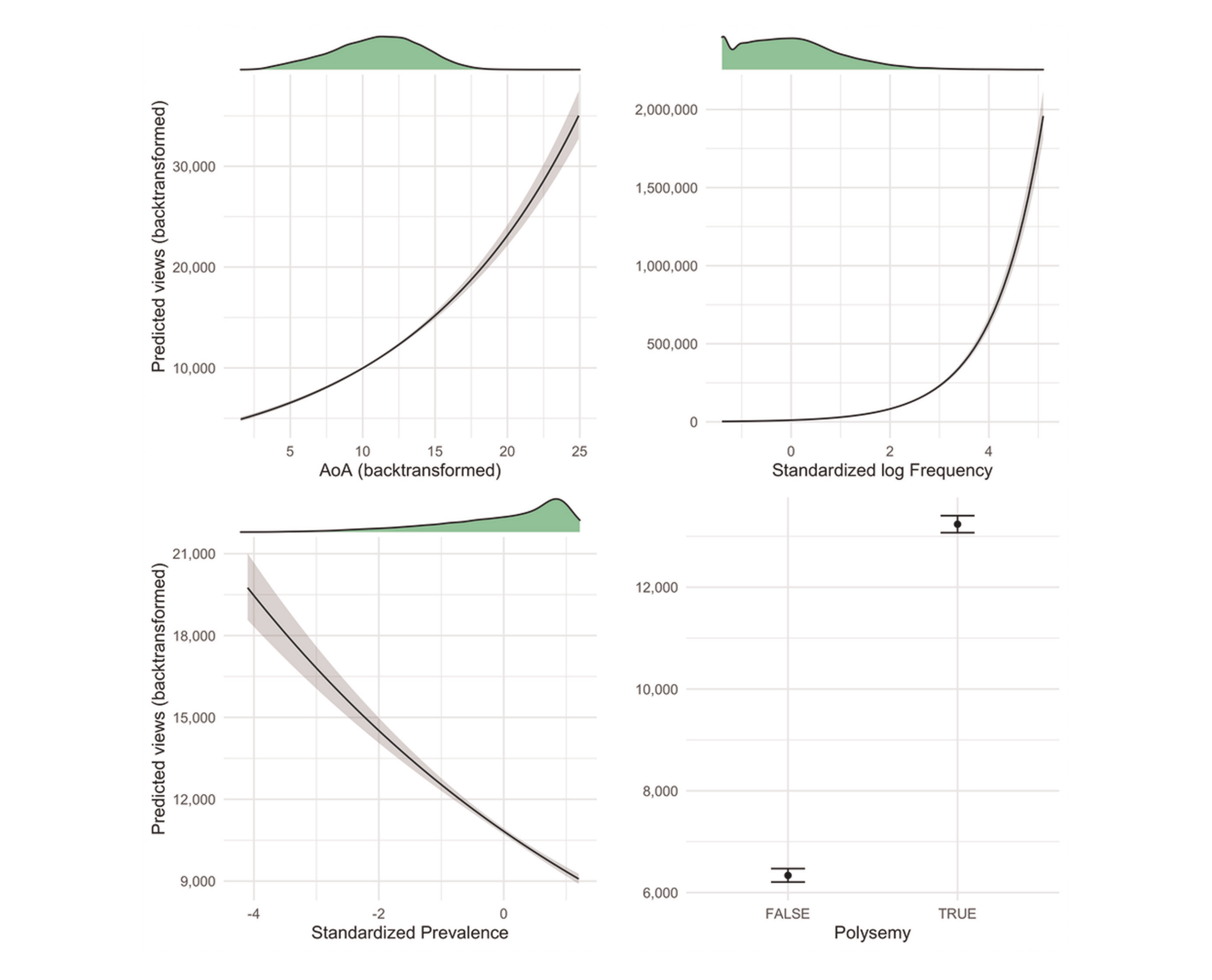

Figure 1. Visualization of the estimated effects of the linear regression model.
Note. Log views (which are predicted by the model) are back-transformed to allow for easier interpretation. For the continuous predictors, we included marginal density plots (green) to give an impression of the distribution of the predictors. Shaded areas/error bars indicate 95% confidence intervals.

This is considerably higher than the entry with the fewest look-ups in the "Verbs" violin ("prewashed," 156 look-ups).

Discussion

The present study confirms the crucial role that lexical frequency plays in driving interest in words for the purpose of dictionary consultation. This finding tallies well with previous studies (De Schryver et al., 2019; Müller-Spitzer et al., 2015), restating that the more frequent words tend to be looked up more often than the less frequent words. We believe a large part of this effect is rather mechanical: corpus frequency represents textual frequency (reflecting the type and proportion of texts represented in the particular corpus), which is also the probability of encountering a given lemma in running text. Now, some dictionary look-ups must be shots in the dark, without specific semantic motivation or well-formed assumptions. Dictionary look-ups that would be so classified would then be expected to reflect the textual frequency of word forms. Thus, the mere higher frequency of occurrence would drive consultation behavior, making it more likely, in a fairly superficial way, for frequent forms to be looked up more often. Simply put, we just cannot help to look up some words just because they are so frequent in the linguistic material that we encounter.

Our regression model also indicates that the other three factors considered also play a role, as suggested by the significance level of these effects and reduction in explained variance (columns 4 and 6 in Table 3). Judging

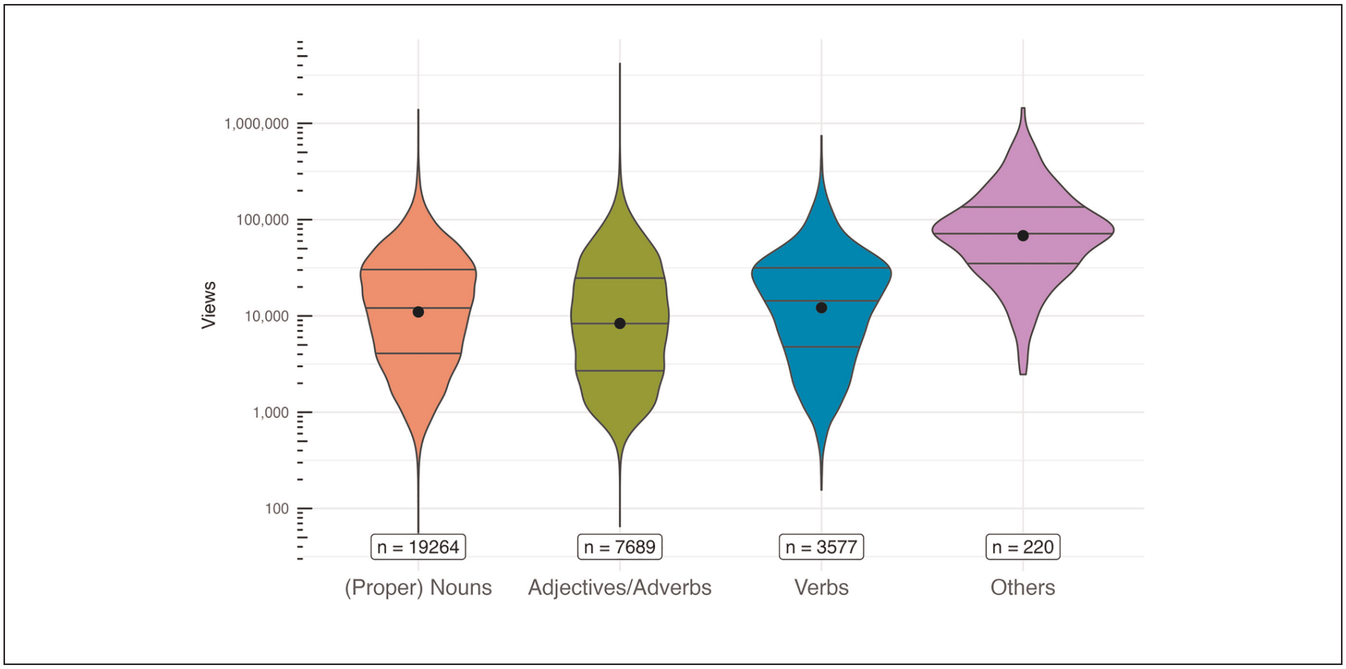

Figure 2. Violin plots for the (log-transformed) look-ups of four part-of-speech categories.
Note. The horizontal lines in each violin represent the first quartile, median, and third quartile, respectively. The point in each violin represents the mean value. The labels show the number of Wiktionary entries in each category.

by the latter parameter (ΔR^2), second in importance would be polysemy.

Polysemous words tend to attract more views. One possible explanation for this effect might be straightforward: encountering words in text, language users pay attention, not just to form, but also meaning. One might even say that meaning is usually the prime goal of communication, with form being a mere vehicle to get at the meaning. This is where the effect of polysemy comes in. A polysemous word could (and, given our results, does) present a greater obstacle when it comes to figuring out the correct meaning of a given word in its context. This might simply be the case because there are several potential "meaning candidates" (some of them might be quite rarely used) from which the correct one must be selected and integrated into the overall meaning representation of the text. This selection process is exactly where dictionaries can help. Compared to that, monosemous words present less of a challenge to the reader/listener. This contrast is reflected in the effect of polysemy in our data. For example, take a word like *school*, whose most common meaning is quite generally known. However, the sense "group of fish" is not so well known, and when this use is encountered in the context of marine life form, users may be puzzled by the idea of fish attending an institution of learning. This type of experience of semantic difficulty may drive dictionary consultation for the less transparent meaning extensions marked as separate senses in the dictionary. This finding again confirms results by Müller-Spitzer et al. (2015) for the German Wiktionary: polysemous words are looked up more often than words with a single meaning.

Next in line in terms of ΔR^2 is age of acquisition (AoA). Our best model indicates a positive effect, which means that words acquired later in life are in general more likely to be looked up. Given that this is an effect adjusted for frequency, we might offer an interpretation of this effect in terms of typical progression of lexical acquisition as well as consultation behavior. A significant part of the core vocabulary of the language is acquired in the early years of life (e.g., Anglin et al., 1993, p. 62, estimate the mean number of main entries in a dictionary known by fifth-graders at around 40,000). Under a typical language acquisition scenario, children would get a good grasp of these words before they start school. On the other hand, pre-school children are not yet literate and would not be expected to use dictionaries such as the English Wiktionary. Conversely, the typical user of Wiktionary already knows (most of) the early-AoA words and does not have to look them up as often as words that they do not know at their current point in life. This might explain why such relatively early-acquisition words are relatively underrepresented in the Wiktionary logs (after correcting for frequency). Note the point about polysemy above, though: rarer senses of early-acquisition words might still drive dictionary consultation.

The final predictor in our model—one with the least impact on the number of views of all—is lexical

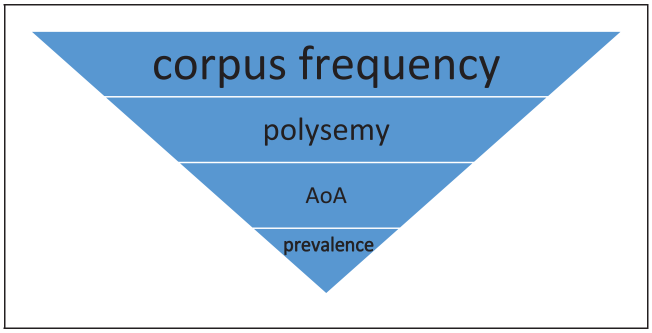

Figure 3. Relative hierarchy of importance of the four predictors of the frequency of dictionary consultation.

prevalence: an indicator of how widespread in the population the knowledge of a given word is. One might expect that if many people know word X, few people would want to look it up. Conversely, if few people know word Y, there would be many who might seek lexical help for it. The direction of the effect in our model agrees with this rationale: words that are known to more people are looked up *less* often (as indicated by the negative estimate).

The final hierarchy of importance that emerges is as follows: frequency > polysemy > AoA > prevalence, and may be rendered graphically as in Figure 3.

Implications for Lexicography

The main implication of our findings for lexicography is one of reassurance: the modern approach to the determination of a lemma list which underlies state-of-the-art corpus-based methodology is essentially correct: attention and effort should be primarily graduated in relation to corpus frequency. Our results tally with those of previous research looking at different types of dictionaries for other languages (De Schryver et al., 2019; Koplenig et al., 2014). However, we have shown that while frequency is a very important predictor of consultation behavior, it is not the only one. While Müller-Spitzer et al. (2015) already showed that short-term effects of social relevance can have impact on look-up behavior, we showed here that there are other long-term variables beside frequency, namely polysemy, age of acquisition, and word prevalence exerting their influence on dictionary users. These factors can (and maybe should) now also be taken into consideration when devising lemma lists or, of course, when expanding on research into dictionary use.

By its very nature, an analysis of web server records cannot shed light on either the context of an instance of dictionary use, or on the personal characteristics of the dictionary user. However, considering the practical application of dictionary-making, specific look-up context as well as idiosyncratic user characteristics cannot be determined at the stage of lexicographic design either, so it seems appropriate to ignore these factors, as in our approach.

Limitations and Future Work

Inevitably in a web-based log-file study with anonymous users, it was not possible to consider the language background and proficiency of the Wiktionary visitors, nor would any other personal characteristics of our dictionary users be known. Likewise, we could not obtain information on such details of the look-up context as the main activity (e.g., was it reading for comprehension, writing, translation work, or perhaps recreational dictionary browsing), or what sort of problem prompted the dictionary look-up. While "know your user" remains a valid principle in lexicography, it is also true that a general-purpose dictionary such as the English Wiktionary attracts a very broad variety of visitors trying to use it for all sorts of purposes. In view of that, the varied log-file data may actually not be such a bad source of information, especially if we consider that the dictionary users whose look-ups we are using are people who came to use the English Wiktionary out of choice, rather than being a captive audience in a controlled experiment.

In the present study, we used a multiple regression model with no interaction terms (though see the Supplemental Material for an interaction model), but other analysis protocols might be employed to yield corroborative or more nuanced results, such as bootstrapped models (or other analyses based on repeated sub-sampling of the dataset).

Another avenue for future research could also include part-of-speech information in the analyses: it might well be that, for example, nouns attract more views than adjectives. At the suggestion of one anonymous reviewer, we tried looking at POS. However, it is still unclear exactly which part-of-speech information should be used. Many entries have more than one part of speech listed; for example, the entry for *angle* includes both noun and verb uses, and it is not clear that the ordering of the POS sections is systematically motivated, nor do we know whether a nominal or verbal use was being looked up. Also, corpus frequency (or other predictors) might affect different parts of speech in different ways.

Furthermore, it might be worth exploring methods from the field of artificial intelligence (AI) or machine learning (ML) to corroborate or extend the present findings, which are based on regression modeling.[6]

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded in part by the National Science Centre, Poland, award 2020/39/B/HS2/00923. For the purpose of Open Access, the authors have applied a CC-BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

ORCID iDs

Robert Lew https://orcid.org/0000-0002-6772-210X
Sascha Wolfer https://orcid.org/0000-0002-8893-8153

Data Availability Statement

The data accomopanying this study as well as the R script used in the analysis are openly available in the OSF repository at https://osf.io/2gejf/.

Supplemental Material

Supplemental material for this article is available online at https://osf.io/2gejf/.

Notes

1. This data was extracted for January, February, March, and April 2022 from https://stats.wikimedia.org/#/en.wiktionary.org/reading/page-views-by-country/normal|table|last-month|(access)~desktop*mobile-app*mobile-web|monthly (last access on 2022-05-10).
2. Relevant data was extracted from https://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population (last access on 2022-05-10).
3. Here we assume that entry structure in the English Wiktionary is more or less consistent. We verified the reliability of the extraction function by spot checking several entries and found no problems.
4. Effect estimates (also called slopes or beta coefficients) from linear regression models give the change of the criterion when the respective predictor changes by a unit of 1. When predictors are not standardized, a unit of 1 is not comparable between the different scales of predictors. Hence, we applied the standardization described above.
5. Others contain abbreviations, articles, conjunctions, contractions, determiners, interjections, letters, numerals, particles, prepositions, and pronouns.
6. In a preliminary analysis, one such technique, binary regression trees, yielded very similar results in terms of importance ranking of our predictor variables.

References

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, *58*(10), i + iii + v-vi + 1–186. https://doi.org/10.2307/1166112

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press. http://lexbib.elex.is/entity/Q12022

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, Supplement*, *10*(3), 89–129. Scopus. https://doi.org/10.1177/1529100610387084

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. Scopus. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

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

De Schryver, G.-M., & Joffe, D. (2004, July 6–10). On how electronic dictionaries are really used. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004*, Lorient, France (*Vol. 1*, pp. 187–196). Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

De Schryver, G.-M., Joffe, D., Joffe, P., & Hillewaert, S. (2006). Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. *Lexikos*, *16*, 67–83.

De Schryver, G.-M., Wolfer, S., & Lew, R. (2019). The relationship between dictionary look-up frequency and corpus frequency revisited: A log-file analysis of a decade of user interaction with a Swahili-English dictionary. *GEMA Online Journal of Language Studies*, *19*(4), 1–27. https://doi.org/10.17576/gema-2019-1904-01

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1103–1123. Scopus. https://doi.org/10.1037/0278-7393.26.5.1103

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. Cambridge Core. https://doi.org/10.1017/S0272263102002024

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468–492. Scopus. https://doi.org/10.1006/jmla.2000.2784

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(2), 267–283. Scopus. https://doi.org/10.1037/0278-7393.24.2.267

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. Scopus. https://doi.org/10.1017/S0305000907008641

Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities*, *34*(1–2), 205–215.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*(5), 684–712. Scopus. https://doi.org/10.1037/0033-2909.131.5.684

Keyes, O., & Lewis, J. (2020). *pageviews: An API Client for Wikimedia Traffic Data*. https://cran.r-project.org/src/contrib/Archive/pageviews/

Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, *31*(2), 91–113.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Koplenig, A., Meyer, P., & Müller-Spitzer, C. (2014). Dictionary users do look up frequent words. A log file analysis. In C. Müller-Spitzer (Ed.), *Using online dictionaries* (pp. 229–249). Walter de Gruyter.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. Scopus. https://doi.org/10.3758/s13428-012-0210-4

Lew, R., & de Schryver, G.-M. (2014). Dictionary users in the digital revolution. *International Journal of Lexicography*, *27*(4), 341–359. https://doi.org/10.1093/ijl/ecu011

Longobardi, E., Rossi-Arnaud, C., Spataro, P., Putnick, D. L., & Bornstein, M. H. (2015). Children's acquisition of nouns and verbs in Italian: Contrasting the roles of frequency and positional salience in maternal language. *Journal of Child Language*, *42*(1), 95–121. Scopus. https://doi.org/10.1017/S0305000913000597

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *Quarterly Journal of Experimental Psychology*, *68*(8), 1623–1642. Scopus. https://doi.org/10.1080/17470218.2014.988735

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. Scopus. https://doi.org/10.1016/j.jml.2016.04.001

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods*, *52*(2), 741–760. Scopus. https://doi.org/10.3758/s13428-019-01272-8

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116–133. https://doi.org/10.1037/0278-7393.21.1.116

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*(6), 705–714. Scopus. https://doi.org/10.3758/BF03202720

Müller-Spitzer, C., Wolfer, S., & Koplenig, A. (2015). Observing online dictionary users: Studies using Wiktionary log files. *International Journal of Lexicography*, *28*(1), 1–26. https://doi.org/10.1093/ijl/ecu029

Navarrete, E., Pastore, M., Valentini, R., & Peressotti, F. (2015). First learned words are not forgotten: Age-of-acquisition effects in the tip-of-the-tongue experience. *Memory & Cognition*, *43*(7), 1085–1103. https://doi.org/10.3758/s13421-015-0525-3

Picard, O., Blondin Massé, A., & Harnad, S. (2010). *Learning word meaning from dictionary definitions: Sensorimotor induction precedes verbal instruction*. Summer Institute on the Origins of Language. Cognitive Sciences Institute. Université du Québec à Montréal.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Sinclair, J. (Ed.). (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. Collins ELT.

Van der Meer, G. (2004, July 6–10). On defining: Polysemy, core meanings, and "great simplicity". In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004*, Lorient, France (*Vol. 2*, pp. 807–815). Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

Verlinde, S., & Binon, J. (2010). Monitoring dictionary use in the electronic age. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the XIV Euralex International Congress* (pp. 1144–1151). Afûk. http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202010/106_Euralex_2010_7_VERLINDE%20BINON_Monitoring%20Dictionary%20Use%20in%20the%20Electronic%20Age.pdf

Verlinde, S., & Selva, T. (2001). *Corpus-based versus intuition-based lexicography: Defining a word list for a French learners' dictionary* [Conference session]. *Proceedings of the 2001 Corpus Linguistics Conference*, Lancaster University (pp. 594–598). http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/verlinde.pdf

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. Scopus. https://doi.org/10.1037/0012-1649.37.2.265

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 205–223. https://doi.org/10.1037/0096-1523.34.1.205

Zipf, G. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.