MorphoLex-FR: A derivational morphological database for 38,840 French words

Hugo Mailhot¹ · Maximiliano A. Wilson² · Joël Macoir² · S. Hélène Deacon³ · Claudia Sánchez-Gutiérrez⁴

Published online: 1 November 2019
© The Psychonomic Society, Inc. 2019

Abstract

Studies on morphological processing in French, as in other languages, have shown disparate results. We argue that a critical and long-overlooked factor that could underlie these diverging results is the methodological differences in the calculation of morphological variables across studies. To address the need for a common morphological database, we present MorphoLex-FR, a sizeable and freely available database with 12 variables for prefixes, roots, and suffixes for the 38,840 words of the French Lexicon Project. MorphoLex-FR constitutes a first step to render future studies addressing morphological processing in French comparable. The procedure we used for morphological segmentation and variable computation is effectively the same as that in MorphoLex, an English morphological database. This will allow for cross-linguistic comparisons of future studies in French and English that will contribute to our understanding of how morphologically complex words are processed. To validate these variables, we explored their influence on lexical decision latencies for morphologically complex nouns in a series of hierarchical regression models. The results indicated that only morphological variables related to the suffix explained lexical decision latencies. The frequency and family size of the suffix exerted facilitatory effects, whereas the percentage of more frequent words in the morphological family of the suffix was inhibitory. Our results are in line with previous studies conducted in French and in English. In conclusion, this database represents a valuable resource for studies on the effect of morphology in visual word processing in French.

Keywords Morphology · Psycholinguistic variables · Lexical decision · Database · Visual word recognition · French

Morphemes are the smallest meaningful units in a language. This means that we can process and understand words that include more than one morpheme on the basis of their constituent morphemes, even if we have never encountered them before (Feldman, Milin, Cho, Moscoso del Prado Martín, & O'Connor, 2015). For example, when reading the invented word *equalism* for the first time we can recognize the form and meaning of its two morphemes: *equal* (its root) and *-ism*

(its affix). The root *equal* appears in several other words such as *equal*, *equality*, or *equalize*. The suffix *-ism* is quite widespread and can be found in words such as *socialism* or *revisionism*, where it confers the concept of an ideology based on the adjective or noun that serves as root (i.e., *social*, *revision*). From this knowledge of the morphemes that form *equalism*, we can easily infer the meaning of *equal* + *-ism* as the ideology that asserts that all humans are equal (Nagy, Carlisle, & Goodwin, 2014). This is an example of derivation. In morphology, there is a distinction between derivation and inflection. Derivation is related to the morphological processes for the creation of new words, whereas inflection is related to the different forms a word can take without changing its meaning (Booij, 2006). For instance, *read*, *reads*, and *reading* are different inflected forms of the same base word *read*, whereas *reader* has a different meaning and lexical category (noun) from *read*, and it is thus a derivation and not an inflection. We present here a comprehensive database of derivational morphological variables for 38,840 French words. We validate this database with experimental research with a visual lexical decision (LD) task.

✉ Maximiliano A. Wilson
maximiliano.wilson@fmed.ulaval.ca

✉ Claudia Sánchez-Gutiérrez
chsanchez@ucdavis.edu

¹ Department of Computer Science, University of California, Davis, CA, USA

² Centre de recherche CERVO et Département de réadaptation, Université Laval, Québec, Canada

³ Dalhousie University, Halifax, Nova Scotia, Canada

⁴ Department of Spanish and Portuguese, University of California, Davis, CA, USA

We do so in part because the importance of morphology in language processing cannot be underestimated. In general, derivational morphology is involved in our ability to understand and create new words such as *manspreading* or *gamification* (Seidenberg & Gonnerman, 2000). As such, morphology influences a plethora of tasks and language processes. For instance, morphology plays a beneficial role in reading acquisition (e.g., D'Alessio, Jaichenco, & Wilson, 2018; Deacon & Francis, 2017), vocabulary learning (Sparks & Deacon, 2015), spelling (Sánchez-Gutiérrez, 2013) and reading comprehension in children (Deacon, Kieffer, & Laroche, 2014). Morphology has also been shown to continue to benefit word processing (i.e., word reading and word recognition) in children with dyslexia who have not yet mastered whole-word processing (e.g., Marcolini, Traficante, Zoccolotti, & Burani, 2011; Suárez-Coalla, Martínez-García, & Cuetos, 2017) and it is related to higher levels of linguistic proficiency in learners of second/foreign languages (Sánchez-Gutiérrez & Hernández-Muñoz, 2018). Also, neurological patients with semantic deficits such as those presenting with the semantic variant of primary progressive aphasia show difficulties in the comprehension and production of morphologically complex words (Auclair-Ouellet, Fossard, Houde, Laforce, & Macoir, 2016; Auclair-Ouellet, Fossard, Laforce, Bier, & Macoir, 2017). The morphological properties of words affect both auditory (e.g., Balling & Baayen, 2008) and visual word recognition (e.g., Oganyan, Wright, & Herschensohn, 2019). Given that the experimental task that we present here is a visual LD task, in what follows we will focus on the behavioral (i.e., time-based responses) effects of morphology on visual word recognition in adults.

Extensive evidence shows that morphemes play an important role in lexical representation and word recognition (see Amenta & Crepaldi, 2012; Feldman & Milin, 2018, for two recent reviews). Nevertheless, the precise way in which morphologically complex words are processed is still under debate. For instance, are all morphologically complex words processed through their constituent morphemes? And, if morphological processing is reserved to a subset of morphologically complex words, what are the psycholinguistic variables that support morphological decomposition over processing as a whole word?

In the literature, whole-word and morphological (i.e., units smaller than whole-words) variables seem to influence visual word processing in different ways. For example, several authors found that morphological decomposition results in faster processing than whole-word processing in the case of low, but not high, frequency words (Alegre & Gordon, 1999; Lehtonen, Niska, Wande, Niemi, & Laine, 2006; Stemmer & MacWhinney, 1986). Notwithstanding, other authors found that all morphologically complex words, not only low-frequency ones, are decomposed into morphemes during lexical processing (McCormick, Brysbaert, & Rastle, 2009;

Rastle & Davis, 2008; Rastle, Davis, & New, 2004). Another finding that has been replicated in some studies is that complex words that include morphemes with a higher summed root frequency (i.e., the sum of the frequencies of all the words that share the same root) are processed faster through morphological decomposition than words with lower summed root frequency (Taft & Ardasinski, 2006). Other studies found that root family size (i.e., the number of words with the same root) modulated morphological decomposition. Words with a larger family size are processed faster through morphological decomposition than are words with smaller families (Balling & Baayen, 2008; Ford, Davis, & Marslen-Wilson, 2010; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2004). However, these findings have both been contested, and there is ongoing debate as to why such divergent results have emerged in the literature (Baayen, Wurm, & Aycock, 2007; Schreuder & Baayen, 1997).

We can evoke several reasons for these divergent results, including different theoretical interpretations of similar results and the lack of uncontested morphological effects (Amenta & Crepaldi, 2012). A recent article suggests another possible and not mutually exclusive explanation for these divergent results; they might emerge from the considerable differences in how morphological variables are calculated across studies (Sánchez-Gutiérrez, Mailhot, Deacon, & Wilson, 2018). For instance, the calculations of summed root frequency or affix frequency are highly dependent on the size of the database on which the calculations are based. Thus, if researchers use a small database to calculate family sizes, they might obtain smaller family size indices than researchers who base these calculations on a broader database. This, of course, will have an impact on results and may actually impede any meaningful comparison across studies. Therefore, we proposed that a way to solve the issue of morphological variables calculation would be to count on a sizeable morphological database in which all the morphological variables are calculated from the same database. With this purpose in mind, we created MorphoLex (Sánchez-Gutiérrez et al., 2018), a comprehensive database of derivational morphological variables for nearly 70,000 English words taken from the complete English Lexicon Project (ELP; Balota et al., 2007). The existence of the database now ensures that future studies on morphological processing in English will be comparable, provided that they use the data from MorphoLex.

Although this is a first step that brings some methodological homogeneity to the field of morphological processing, the benefits of MorphoLex are limited to studies that focus on the English language. This is restrictive, as we know that languages differ significantly in relation to their morphological structure. For example, as compared with French, English relies consistently on compound structures, in which a root (e.g., *brush*) is added to another root (e.g., *tooth*) to create a new word (e.g., *toothbrush*). Zero derivation is also typical in

English, whereby, for example, the adjective *clear* does not need to get an affix appended to it in order to be used as a verb in sentences such as “We need to *clear* the route.” The strong reliance on both compounding and zero derivation is characteristic of the English language and is unusual in other languages. For example, there are very few cases of zero derivation in French, which relies mostly on derivational processes. These typological differences might affect how morphologically complex words are processed in both English and French. Therefore, the aim of the present study is to present MorphoLexFR, a sizeable French database of derivational morphology based on the 38,840 words of the French Lexicon Project (FLP; Ferrand et al., 2010). We used similar procedures for the segmentation and calculations of morphological variables as those used in English for MorphoLex.

Morphological processing in French

MorphoLexFR would be useful in part because studies of morphological processing in French show the same inconsistencies as in English (Sánchez-Gutiérrez et al., 2018). For example, the role of summed root frequency and its interaction with family size and whole-word frequency is still a matter of debate in French. Although there is extensive evidence of the impact of root frequency on morphological processing in French, this effect seems to be modulated by other variables, or by morphological word structures—that is, suffixed versus prefixed words (Colé, Beauvillain, & Segui, 1989). This results in a landscape of complex interactions and sometimes contradictory results among studies.

Meunier and Segui (1999) showed that root summed frequency modulated the effect of whole-word frequency in LD latencies of suffixed words. More importantly, they found that when they controlled for the number of words with a higher whole-word frequency than the target in the morphological family (i.e., the percentage or proportion of more frequent words in the morphological family), then root summed frequency affected words of both low and high whole-word frequency. This shows that decision times for a given morphologically complex word depend on the position of this word in its morphological family, which is an extremely relevant finding. Similarly, Colé et al. (1989) found that whole-word frequency modulated the effects of root frequency and morphological root family size in LD latencies. However, this effect was only found for suffixed—but not prefixed—words. Similarly, Beauvillain (1996), using an eyetracking paradigm and a semantic categorization task, showed a different processing for suffixed and prefixed word. Nevertheless, unlike Colé et al., she found that cumulative root frequency—but not whole-word frequency—differentially modulated fixation duration for suffixed and prefixed words. Indeed, cumulative

root frequency influenced the first fixation in suffixed words and the second fixation in prefixed words.

Morphological processing has been widely studied by means of the manipulation of pseudoword structure (i.e., invented words with a real morphological structure vs. pseudowords with no morphological structure) to assess the effects of morphological variables. In a now classic study, Longtin and Meunier (2005) showed that pseudowords that had a transparent morphological structure (i.e., both the root and the suffix are easily identifiable) facilitated the recognition of their roots in a masked priming paradigm. Thus, *rapidifier*, which is a pseudoword but includes a transparent root, *rapide* [fast], and a recognizable suffix, *ifier* [ify], facilitated the processing of its root, *rapide*. Interestingly, this was true for pseudowords that were interpretable semantically, such as *rapidifier*, and pseudowords that were noninterpretable, such as *sportation*. The latter example is noninterpretable because the French language does not allow adding the suffix *-ation* to nouns, because this suffix is used to derive nouns from verbs. Longtin and Meunier’s finding suggests that pseudowords that include a recognizable root and affix are decomposed into their constituent morphemes during processing, even when the resulting pseudoword itself cannot be semantically interpreted. These results were replicated in a follow-up study, using a cross-modal priming paradigm (Meunier & Longtin, 2007).

Morris, Porter, Grainger, and Holcomb (2011) also found morphological priming between pseudowords that included a real suffix and their roots (e.g., *rapidifier* and *rapide*), but this effect was also observed in pseudowords with word endings that were not morphological (e.g., *rapiduit*, where *-uit* is not a French suffix). In this context, Beyersmann, Casalis, Ziegler, and Grainger (2015) performed a masked priming LD experiment manipulating the presence of suffixes and the lexicality of the primes to obtain three types of primes (i.e., suffixed words, suffixed nonwords, and nonsuffixed nonwords). Importantly, the participants in this study were divided into two groups depending on their language proficiency (high vs. low proficiency). Beyersmann et al. (2015) found that language proficiency modulated the pattern of results. Participants in the high-proficiency group showed comparable priming effects for the three types of primes, with no difference between pseudowords that included a real suffix and those that included other types of nonmorphological endings. Conversely, participants in the low-proficiency group showed larger priming effects for the suffixed conditions (i.e., suffixed words, suffixed nonwords) and reduced facilitation for the nonsuffixed condition. Beyersmann, Cavalli, Casalis, and Colé (2016) replicated the results found by Beyersmann et al. (2015), by showing that priming was also modulated by reading proficiency. Highly proficient individuals showed priming effects even in the non-suffixed nonwords prime condition. Taken together, the results of the high-proficiency

group suggest that embedded stems are salient enough to be activated independently of whether they are in combination with a real affix or a nonaffix. These studies, while extremely relevant, mostly used pseudowords. Little research has studied priming experiments with real words, both as primes and targets, in French. We argue that this is partly due to the lack of reliable data on the psycholinguistic variables that are relevant for morphological processing.

Another issue in the literature is the time course of morphological processing (Diependaele, Sandra, & Grainger, 2005). Indeed, several authors have found that morphological processing is based on semantics (Marslen-Wilson, Tyler, Waksler, & Older, 1994), whereas others have found pre-semantic morphological processing (Longtin, Segui, & Halle, 2003). In French, Beyersmann, Iakimova, Ziegler, and Colé (2014) conducted an event-related potential (ERP) study in which they found a partially overlapped activation for morphological and semantic priming effects for LD latencies. However, morphological priming showed a different parietal signature and was present from an earlier time window (100–250 ms), as compared to semantic priming. Their results show an early morphological decomposition effect that goes beyond semantic, and even orthographic, effects. Cavalli et al. (2017) reproduced this early morphological effect (100–200 ms) in adults with developmental dyslexia with magnetoencephalography (MEG) as imaging technique. They also used a primed LD task that contrasted morphological, semantic, and orthographic relationships between primes and targets. Interestingly, they found that adults with developmental dyslexia relied on morphological information more than normal adult readers. This is similar to the results of the low-proficiency group in Beyersmann et al. (2015), which showed a larger morphological reliance (i.e., larger priming effects for suffixed words and nonwords and reduced for nonsuffixed nonwords) than in the high-proficiency group. In general, these studies suggest that low-proficiency readers, as well as dyslexic adults, rely more on morphological information than more proficient participants.

In conclusion, this literature review on morphological processing in French shows that the tasks (e.g., word reading, LD, semantic categorization, etc.), the experimental paradigms (e.g., priming) and the imaging techniques (ERPs, MEG, etc.) are quite varied and heterogeneous among studies. These methodological differences might partially explain the contradictory results found in these studies. Notwithstanding, as we mentioned before, another critical and long-overlooked factor could underlie these diverging results—namely, methodological differences in calculation of the morphological variables in these studies (Sánchez-Gutiérrez et al., 2018). Thus, it is reasonable to argue that a way to overcome this limitation and to make reliable comparisons across studies would be to count on shared morphological databases.

Resources for investigating morphology in French

To the best of our knowledge, five morphological resources currently exist in French, none of which captures the most studied morphological variables (e.g., morpheme frequency, family size, morpheme length, and the position of the word in its family). Morphalou is a morphological dictionary for French words (Romary, Salmon-Alt, & Francopoulo, 2004), which captures on inflectional, but not derivational morphology. As such, it is not adequately comprehensive. The MuLeXFoR database provides information on French, English and Italian affixes (Cartoni & Lefer, 2010). It works in a dictionary-like fashion and provides general information on the rules and uses of a particular affix. It does not, however, provide frequency-based information for morphological variables. The database POLYMOTS provides information on derivational morphology (Gala & Rey, 2008). It includes data on the number of words that belong to the same family (i.e., morphological family size) and the frequency of appearance of a given affix. However, no information is provided about the corpus on which the calculations were based or the rules used to organize the families (e.g., did they include each inflectional form as a different member of the family?). DériF (Namer, 2003, 2013) is a morphological analyzer that decomposes French derived and compound words into their bases and affixes and provides a semantic description of these morphological units. It is useful in enabling segmentation, but it does not provide frequency based information for morphological variables. Finally, the database Manulex-morpho (Peereman, Sprenger-Charolles, & Messaoud-Galusi, 2013) provides morphological information, primarily inflectional, for frequent words encountered by children. It is based on a subset of about 10,000 words of the Manulex database (Lété, Sprenger-Charolles, & Colé, 2004). The words in the Manulex database have been selected from a wide selection of French elementary school texts used from first to fifth grades. It is not clear that this database is applicable to research with adults or on derived morphemes.

In sum, none of the currently available resources in French offers indices for all the variables known to modulate morphological processing. It is noteworthy that only POLYMOTS provides quantitative information on one of the relevant psycholinguistic morphological variables (i.e., morphological family size). All other resources focus more on the morphological description of the words they contain, rather than on morphological variables (Cartoni & Lefer, 2010; Namer, 2003; Romary et al., 2004).

This situation calls for a sizeable morphological database that offers reliable indices of the most studied morphological variables in the literature for a significant number of French words. MorphoLex-FR, the database that we present here, aims to address this need. We created a database of

derivational morphology based on the 38,840 words of the FLP (Ferrand et al., 2010). We chose FLP as a base because this is the largest freely available database in French that includes a collection of LD times. Critically, the FLP followed the same design as the ELP (Balota et al., 2007). This means that the FLP and the ELP are comparable, which renders MorphoLex-FR (French) and MorphoLex (English), based on the ELP, also comparable. This is an important feature, because if the values of morphological variables are taken from equivalent databases, then cross-linguistic comparison, at least in French and English, is possible. MorphoLex-FR offers data on four morphological variables: (1) affix and root frequency, (2) affix and root family size, (3) proportion of words more frequent in the family size (PFMF) for affixes and roots, and (4) affix length. We calculated all of the variables in a similar way to that used in MorphoLex. In the following sections, we present the segmentation method, the new morphological variables calculated for the FLP and the distributional characteristics of this freely available database.

Word segmentation method

We segmented into morphemes the 32,185 morphologically complex words of the FLP. To that end, we used the same codes as in Sánchez-Gutiérrez et al. (2018): < > for prefixes, > > for suffixes, () for roots, and { } for the largest free lexical units in the word. A root is a single nonaffixal morpheme of a word that makes the most concrete and distinctive contribution to the meanings of this word (Carstairs-McCarthy, 2006). For example, *lire* [to read] in *illisible* [illegible], whose form in that particular case is *lis*. The concept of root can be distinguished from those of base and stem. A base is the word from which another word is derived (e.g., *illisible* [illegible] is derived from the word *lisible* [legible]) (Marslen-Wilson, 2006). A stem is what is left of a word once suffixes are removed to reduce morphologically and semantically related words to their common stem (Paice, 2006). This stem does not have to be an actual word (e.g., *illis* in *illisible* [illegible-singular] and *illisibles* [illegible-plural]). Here we focus on roots only. For example, the word *dénégation* [denial] is segmented as <dé<{(nég)>ation>}. We added the square brackets [] to this notation scheme to indicate verbal derivational suffixes, such as the verbal infinitive *-er* in *allonger* [to extend], which gives <al<{(long)}>[er]>. This addition helped us to distinguish between verbal and nonverbal homographic morphemes and to count all verbal derivational suffixes as one category. We treated compounds as containing sequences of roots. For example, the compound word *lave-vaisselle* [dishwasher] was segmented as {(lave)}-{(vaisselle)}.

The decision of whether to place a morpheme boundary was made on the basis of two criteria. The first was semantic:

A word can be segmented only into meaningful units—that is, in a way in which the meaning of the whole word is the result of the combination of the parts in contemporary French. For instance, words like *déduire* [to deduce], *conduire* [to drive], and *produire* [to produce] have long lost the semantic relatedness they originally had in Latin. That is why we considered them here as monomorphemic. Conversely, a word like *incomparable* [incomparable] and *allonger* [to extend] are both segmented into one prefix (i.e., *in-* and *a-*, respectively), one root (i.e., *compar* and *long*), and one suffix (i.e., *-able* and *-er*). Judgments of semantic relatedness were made on the basis of available information from etymological resources and the knowledge of the authors as native and fluent speakers of the French language. First, we observed whether the original meaning of a given word was preserved in such a way that its meaning could still be derived synchronically from its parts in present day French. Afterward, and for the sake of consistency, we observed whether the candidate morpheme was found in other words with a similar meaning. We then identified this morpheme as a meaningful unit and we segmented it accordingly.

Our decision to place a morphemic boundary only when the derived word could be segmented into meaningful smaller units is consistent with the dominant conceptualization of morphemes, as providing an important element of structure to otherwise arbitrary mappings between word forms and their meanings (Hockett, 1958; Quémar, Casalis, & Colé, 2011). Within this conceptualization, the meanings of roots are largely preserved in their derivations (e.g., *bake*, *bakery*, *baker*), and the meaning of the resulting derived words are highly predictable (e.g., the words *nicer*, *bigger* and *smaller* are related in meaning to the words *nice*, *big*, and *small*). On the basis of this conceptualization, many theoretical models of morphological processing propose that only semantically transparent complex words (i.e., the meaning of the complex word can be derived from the meaning of its constituents) share lexical representations with their morphemic constituents (Giraud & Grainger, 2000; Marslen-Wilson et al., 1994). By contrast, the meaning of semantically opaque complex words is unrelated to the meanings of their constituents. In this study, we adopted what we consider to be the most objective method to segment morphologically derived words—that is, semantically based segmentation.

The second criterion was paradigmatic: a morpheme is only recognized as such if it (1) appears in more than one context or in other words and (2) can be identified as part of a morphological system. For example, the word *arbuste* [bush], even though it is semantically related to words like *arbre* [tree] and *arboricole* [arborescent], was not segmented here into a root and a suffix, since *-uste*, as a suffix, is found in no other French word. Since MorphoLex-FR is a derivational database, we ignored inflectional morphemes and thus considered as allomorphs the gender and number variations of a root or a

suffix (e.g., *arbustes* [bushes] was considered allomorph of *arbuste* [bush]). Indeed, as in the English derivational database (Sánchez-Gutiérrez et al., 2018), the words in MorphoLex-FR are classified according to their specific prefix–root–suffix (PRS) signature. Had we taken into consideration inflectional suffixes, it would have changed the signatures of words that, from a derivational point of view, are the same. For instance, a word like “pensive” [pensive-female-singular] would have been classified as having a 0–1–1 PRS ($\{(\text{pens})\}>\text{ive}>$), whereas its plural form “pensives” [pensive-female-plural] would have had a 0–1–2 PRS ($\{(\text{pens})\}>\text{ive}>\text{s}>$).

The classification of a morpheme as either a prefix, a root, or a suffix in the database was generally straightforward. However, some neoclassical compounds posed additional difficulties. For example, *thermo-* in *thermodynamique* [thermodynamics] appears as a prefix at the beginning of the word, but in *thermique* [thermal] it is clearly a root. In such cases, we considered that words such as *thermodynamique* contain two roots. A complex word such as *thermodynamique* [thermodynamics] results from a word formation process called “compounding,” consisting in the linking of two lexical units. Generally, these lexical units are not phonologically or orthographically modified by the composition. Moreover, many of these complex forms are neoclassical compounds, usually words of Latin or Greek origin, belonging to technical vocabulary (Apothélos, 2002; Corbin, 2004). In French, the root *thermo* is used alone, as a free morpheme, in one word (*thermos*) and is found in 13 other complex words (e.g., *thermodynamique* [thermodynamics], *thermometer* [thermometer], etc.). The root *dynamique* is used alone, as a free morpheme, in one word and is found in five other complex words (e.g., *thermodynamique* [thermodynamics], *aérodynamique* [aerodynamics], etc.). Therefore, such roots could be considered as affixoids (i.e., morphemes that function as suffixes and have corresponding lexemes). Affixoids are one of the results of grammaticalization and contain words that are becoming grammatical morphemes. As was stated by Booij (2005), there is no clear boundary between affixal derivation and compounding. Without a theoretical model of morphology considering these particular constituents, the notion of an affixoid should be seen as a convenient and provisional classificatory term. Given these considerations, we decided to segment these complex words as two roots. We thus considered all classical (i.e., Greco-Latin) morphemes as roots (Bauer & Nation, 1993) except those that indicated either a position (e.g., *pré-*, *sou-*, *trans-*, *supra-*), a negation (e.g., *non-*), or a quantity (e.g., *ultra-*, *méga-*, *maxi-*, *bi-*, *tri-*, *déca-*), when such morphemes were not the only potential roots of the words they were in. Thus, we segmented *uni-* in *unidirectionnel* [unidirectional] as a prefix, because the first morpheme, *unité* [unity], has a meaning of quantity.

Furthermore, since *uni-* in *unité* [unity] has no other valid candidate roots, we segmented it as $\{(\text{un})\}>\text{ité}>$. This example shows that, in a few cases, the same morpheme can belong to different morphemic categories (*uni* here is both a prefix and a root).

The same morpheme can have many allomorphs—that is, surface forms. To ensure that we counted the orthographical variations of the same morpheme together, we mapped each set of allomorphs onto a canonical form. For example the suffixes *-able(s)* (as in *aimable* [kind]) and *-ible(s)* (as in *corruptible* [corruptible]) were both annotated as $>\text{able}>$. Thus, we considered the frequency counts of *-able*, *-ables*, *-ible*, and *-ibles* to calculate the frequency of the canonical morpheme *-able*. Both suffixes (*-able(s)* and *-ible(s)*) have the same semantic value (i.e., in French, *aimable*, “that can be loved,” *corruptible*, “that can be corrupted”). Consistent with our semantically based segmentation choice, we privileged the canonical form that represents the shared semantic value of all the members of the morphological family, irrespective of the specific allomorphs. This segmentation decision prioritizes lemma calculations; for instance, we arrive at a lemma calculation for *-able*. It also has the consequence of not allowing the calculation of each allomorph of the suffix separately (e.g., *-able* vs. *-ble*). This segmentation decision also provides a single length of morpheme calculation for all allomorphs based on the lemma. For instance, the allomorph *-able* as in *aimable* [kind] with four letters and its allomorph *-ibles* as in *corruptibles* [corruptible-plural] with five letters will both compute for the length of the canonical morpheme *-able* with four letters of length. For monomorphemic words, this mapping amounts to lemmatization. To focus on derivational morphology only, we used the same principle and mapped all inflectional parts of verbal derivational suffixes as [VB]. For instance, we segmented both *travailleraient* [(they) would work] and *travaillions* [(we) worked] as $\{(\text{travail})\}>[\text{VB}]>$, and *naturaliser* [naturalize] as $\{(\text{nature})\}>\text{el}>[\text{VB}]>$. In this way, we counted together verbal allomorphic suffixes but were able to distinguish between different verbal suffixal morphemes, as long as they were not homographic.

Homographs were not differentiated in the database if they pertained to the same morphemic category (i.e., prefix, root, suffix). For example, we considered one and the same the prefix *in-*, whether it means “inward movement,” as in *incarcération* [incarceration], or negation, as in *incalculable* [incalculable]. However, our database does differentiate morphological homographs when they are encountered in different positions (i.e., suffixes and prefixes). For example, we distinguished the prefix *at-* in *attendrir* [to soften] from the suffix *-at* in *attentat* [attack]. In the database, we did not mark in any way allomorphs that were orthographically identical (e.g., *lent* [slow] vs. *lent-eur* [slowness]) but phonologically distinct (/lã/ and /lât/, respectively).

The new morphological variables

On the basis of the literature review and the segmentations reported in the previous sections, we calculated four new variables for each morpheme. We describe these variables in the following paragraphs.

Morphological family size This is the number of word types in which a given morpheme is a constituent (Baayen, Feldman, & Schreuder, 2006). We calculated the family size of a morpheme by counting all its types in the FLP database. For instance, in {*retenter*, *refaire*, *faire-part*, *redemander*}, the prefix *re-* has a morphological family size of 3 {*retenter*, *refaire*, *redemander*}, whereas the root *faire* has a morphological family size of 2 {*refaire*, *faire-part*}.

Summed token frequency This is the summed frequencies of all members in the morphological family of a morpheme (Sánchez-Gutiérrez et al., 2018). Thus, following the previous example, the frequency of the root *faire* would be the result of adding the frequency of the word *refaire* and that of the word *faire-part*. We used the frequency count *cfreqmovies*, based on films subtitles (New, Brysbaert, Véronis, & Pallier, 2007), taken from the Lexique database (lexique.org; New, Pallier, Brysbaert, & Ferrand, 2004; New, Pallier, & Ferrand, 2005).

Proportion of other words in the family that are more frequent (PFMF) For each morpheme of each word, we computed the PFMF value as the total words in the morphological family, as a fraction of 1, that are more frequent than the given word. The resulting values range from 0 to 1, where 0 means that no word in the family is more frequent, and 1 means that all words in the family are more frequent. For example, *feuille*, *feuillu*, *feuilleter*, and so forth, share the same root (i.e., *feuille* [leaf/sheet]) and thus have an identical family size (of 11, in this case). However, they have different PFMFs: *feuille* does not have any more frequent competitors in the family, and thus it has a PFMF of 0; *feuillu* has 9 words out of 11 that are more frequent in the family, which results in a PFMF of .82; and *feuilleter* has 5 words out of 11 that are more frequent in the family, and thus has a PFMF of .45.

Morpheme length This variable indicates the number of letters in a particular morpheme. We calculated morpheme length for the canonic form of each morpheme and not for each of its allomorphs. For example, *-ion* always has a length of 3, even when it appears as *-tion* or *-ation*.

The database

Each word in the database was tagged with a specific PRS signature. This means that words that include one suffix and

one root, but no prefix, share a 0–1–1 PRS signature (i.e., 0 for the number of prefixes, 1 for the number of roots, and 1 for the number of suffixes), whereas words with two roots and a prefix will be tagged as 1–2–0 (i.e., 1 for the number of prefixes, 2 for the number of roots, and 0 for the number of suffixes). For example, the word *stagiaire* [intern] has a 0–1–1 PRS, because it includes no prefix, one root (i.e., *stage*), and one suffix (i.e., *-aire*). The database is presented in an Excel file that is freely available at the following address: <https://github.com/hugomailhot/MorphoLex-FR>.

Each PRS signature appears in different sheets that are titled with the corresponding PRS signature. This allows researchers to directly access any specific subset of words, depending on their morphological structure. The first page offers a list of all the variables and their corresponding headers, in order to facilitate the queries. For each morpheme on Sheets 2–17, the above-mentioned morphological variables are provided in columns titled with the name of the variable, preceded by ROOT, PREF, or SUFF and a number (e.g., ROOT1, PREF2, etc.). That number indicates the location of the morpheme in the word. For example, ROOT1 will be the first root in the word, and PREF2 will be the second prefix in the word. Sheets 18, 19, and 20 list all the prefixes, suffixes, and roots in the database, respectively. This allows users to obtain specific information about each morpheme, independent of the words in which it is embedded. We expect the morpheme lists to be particularly useful when creating morphologically complex pseudowords.

Descriptive analyses for the database

In this section we report some of the main findings derived from the database. Overall, 41% of the words in the database were monomorphemic (PRS 0_1_0), whereas 59% of the words were morphologically complex—that is to say, had at least one prefix, one suffix, or more than one root (i.e., all other PRSs). This resembles fairly closely the distribution of morphologically complex and simple words in French in Rey-Debove (1984). Table 1 shows the distribution of the different PRSs, in decreasing order according to the percentage of words in the database. The most prevalent morphologically complex type of words in the database is words with one suffix (PRS 0_1_1). This type of word alone represents nearly 37% of the database and closely follows the number of morphologically simple words. The second and third most prevalent types of morphologically complex words are those with one prefix and one suffix (PRS 1_1_1; 8% of the total database) and those with one prefix (PRS 1_1_0; 5%). The fourth most common complex type of word are those with two suffixes (PRS 0_1_2; 4%). All the other types of complex words together represented a mere 5% of the total database.

Table 1 Summary of the database as a function of the PRS signatures

PRS signature	Number of words	Percentage	Example	Segmentation	English translation
0_1_0	6655	41.26	abricot	(abricot)	apricot
0_1_1	5962	36.96	absence	(absent)>nce>	absence
1_1_1	1367	8.47	injustice	<in<(juste)>ice>	injustice
1_1_0	752	4.66	collatérale	<co<(latéral)	collateral
0_1_2	668	4.14	psychiatrie	(psycho)>iatr>>ie>	psychiatry
0_2_0	313	1.94	antidote	(anti)(dote)	antidote
0_2_1	177	1.10	agriculture	(agricole)(cult)>ure>	agriculture
1_1_2	97	0.60	immortalité	<in<(mort)>el>>ité>	immortality
2_1_1	68	0.42	coalition	<co<<a<(lier)>ion>	coalition
2_1_0	27	0.17	renfort	<re<<en<(fort)	backup
0_1_3	22	0.14	professionnalisme	(professer)>ion>>el>>isme>	professionalism
0_2_2	7	0.04	radioactivité	(radio)(acte)>if>>ité>	radioactivity
0_3_0	6	0.04	paléontologie	(paléo)(onto)(logue/ie)	paleontology
2_1_2	5	0.03	désintoxication	<dé<<in<(tox)>ique>>ion>	detoxification
1_1_3	2	0.01	infinitésimal	<in<(fin)>ité>>ésim>>al>	infinitesimal
1_2_1	2	0.01	invraisemblance	<in<(vrai)(sembler)>nce>	unlikeliness
1_2_0	1	0.01	analphabète	<a<(alpha)(beta)	illiterate

PRS, prefix–root–suffix; Number of words, number of words in the database; Percentage, percentage of words in the database for a given PRS signature

For the sake of simplicity, in what follows we will only focus on the three most common types of morphological words in the database (i.e., PRS 0_1_1, 1_1_1, and 1_1_0) and compare them to morphologically simple words (PRS 0_1_0). To simplify the scale, the values for family size and summed token frequency are log-transformed in the figures. PFMF is presented as a proportion (i.e., from 0 to 1), and morpheme length in number of letters.

Morphological family size As is shown in Fig. 1, morphologically complex words tend to have larger root morphological families than morphologically simple words (PRS 0_1_0). Also, when it comes to affixation, suffixed words show larger

morphological families than do roots. Suffixed words (PRS 1_1_1 and 0_1_1) have larger families than prefixed words. This indicates that suffixes belong to larger morphological families and thus are more productive than prefixes and roots.

Summed token frequency Figure 2 shows that the roots in morphologically complex words are, on average, of higher summed frequencies than simple words. Likewise, prefixed words (i.e., PRS 1_1_0 and 1_1_1) have higher token frequencies than either suffixed (PRS 0_1_1) or simple words (PRS 0_1_0). Additionally, the morphological family members of the suffixes are more frequent than the family members of prefixes.

Fig. 1 Log-transformed mean family size as a function of prefix–root–suffix (PRS) signature for prefixes, roots, and suffixes

Fig. 2 Log-transformed summed token frequency as a function of prefix–root–suffix (PRS) signature for prefixes, roots, and suffixes

PFMF As expected, morphologically simple words, which act as roots in all derived and compound words in their family, are more frequent words in their morphological families than morphologically complex words (PRS 1_1_0, 1_1_1, and 0_1_1) (see Fig. 3). Also, the suffixes in these PRS signatures (i.e., 1_1_1 and 0_1_1) are associated with less frequent words in their families than are prefixes.

Morpheme length Figure 4 shows the mean morpheme lengths of the four PRS signatures presented here. The lengths of the canonic forms of the roots for the four PRS signatures are comparable, though roots seem slightly larger for morphologically simple words. Likewise, suffixes are larger than prefixes in terms of their number of letters.

Influence of the morphological variables on lexical decision latencies

We report here on a set of analyses designed to explore the influence of the newly developed French morphological variables (i.e., frequency, family size, PFMF, and length) to LD reaction times (RTs). This offers a validation of these variables. For the sake of comparability with the English morphological database MorphoLex (Sánchez-Gutiérrez et al., 2018), we focused here on nouns containing one root and one suffix (i.e., a 0–1–1 PRS signature). We extracted the RTs for these word types from the FLP (Ferrand et al., 2010). We then entered the values of the morphological variables and other relevant psycholinguistic variables (e.g., frequency, imageability, etc.) as predictors in a series of hierarchical regression models.

Fig. 3 Proportion of other words in the family that are more frequent (PFMF) as a function of prefix–root–suffix (PRS) signature for prefixes, roots, and suffixes

Fig. 4 Mean morpheme length (in number of letters) as a function of prefix–root–suffix (PRS) signature for prefixes, roots, and suffixes

Method

Materials

Out of the 5,962 words of the FLP that had one root and one suffix (i.e., PRS 0–1–1), we selected the 917 nouns in common with the available French database for imageability (Desrochers & Thompson, 2009). Only ten had values for age of acquisition in French (Lachaud, 2007), and none of the 917 had values for concreteness (Bonin, Méot, & Bugaiska, 2018). It is noteworthy that both databases contain mostly monosyllabic words (only 137 words are bisyllabic).

This reduces the possibility of finding morphologically complex French words such as the ones we used in the present study. The final set of stimuli for the study was thus composed of the RTs to 917 nouns. Table 2 shows the summary statistics for all the variables for these nouns.

We obtained the mean LD RTs for the 917 nouns from the FLP (Ferrand et al., 2010). To conduct analyses comparable to those reported in Sánchez-Gutiérrez et al. (2018), we also extracted the values for the following psycholinguistic variables for the 917 nouns from the database Lexique 3 (lexique.org; New et al., 2004; New et al., 2005): word length in number of letters (N-letters), word length in number of

Table 2 Summary statistics for all variables used in the lexical decision study

	Mean	Standard Deviation	Minimum	Maximum	Skewness	Tolerance	VIF
Length	2.92	0.86	1	6	0.5	0.39	2.57
log Frequency	0.85	0.42	0.03	2.14	0.44	0.64	1.57
log NSize	0.18	0.24	0.00	1.11	1.11	0.59	1.7
OLD20	2.43	0.54	1.15	4.75	0.63	0.34	2.97
Imageability	3.82	1.28	1.64	6.9	0.66	0.75	1.34
Root length	6.16	1.77	2	12	0.30	0.55	1.83
log Root Frequency	1.59	0.65	0.004	4.32	0.24	0.35	2.86
log Root Family Size	0.63	0.28	0.00	1.49	0.24	0.53	1.89
Root PFMF	31.52	21.88	0.00	87.5	0.1	0.58	1.73
Suffix length	3.14	0.83	1	7	1.75	0.94	1.06
log Suffix Frequency	2.91	0.71	0.06	4.40	-1.13	0.28	3.52
Suffix Family Size	367.52	329.43	1	1600	0.84	0.29	3.36
Suffix PFMF	26.37	20.92	0.00	94.61	0.79	0.44	2.27
RTs	698.65	67.63	571.91	1044.09	0.92	N/A	N/A

Length, length in syllables; log Frequency, log-transformed frequency; log NSize, log-transformed orthographic neighborhood size; OLD20, orthographic Levenshtein distance 20; Root length, root length in letters; log Root Frequency, log-transformed root frequency; log Root Family Size, log-transformed root family size; Root PFMF, percentage of more frequent words in the morphological family of the root; Suffix length, suffix length in letters; log Suffix Frequency, log-transformed suffix frequency; Suffix PFMF, percentage of more frequent words in the morphological family of the suffix; RTs, LD latencies. Tolerance and variance inflation factor (VIF) are estimations of multicollinearity

Table 3 Correlations between all the variables used as predictors (and the dependent variable RTs) in the lexical decision task

	Length	log Frequency	log NSize	OLD20	Imageability	Root length	log Root Frequency	log Root Size	Root Family Size	Root PFMF	Suffix length	log Suffix Frequency	Suffix Family Size	Suffix PFMF	RTs
Length	1														
log Frequency	-.022	1													
log NSize	-.406**	.004	1												
OLD20	.642**	-.061		1											
Imageability	-.394**	-.056	.150**	-.246**	1										
Root length	.598**	.007	-.297**	.512**	-.266**	1									
log Root Frequency	-.232**	.363**	.149**	-.237**	.080*	.376**	1								
log Root Family Size	-.254**	.185**	.134**	-.268**	.028	-.216**	.635**	1							
Root PFMF	.025	-.290**	-.006	.006	-.096**	-.026	.186**	.065	1						
Suffix length	.066*	-.062	-.105**	.135**	.000	.066*	.030	.031	.044	1					
log Suffix Frequency	.387**	.116**	-.050	.047	-.371**	-.174**	-.077*	-.091**	-.087**	.033	1				
Suffix Family Size	.430**		.074*	-.101**	.084*	-.373**	.367**	-.073*	-.087**	-.032	.089**	.818**	1		
Suffix PFMF	.053	-.520**	-.056	.103**	-.105**	-.044	-.431**	-.157**	.406**	-.041	.049		-.011	1	
RTs	.179**	-.401**	-.115**	.253**	-.109**	.037	-.332**	-.183**	.138**	-.037	-.155**		-.082*	.420**	1

Length, length in syllables; log Frequency, log-transformed frequency; log NSize, log-transformed orthographic neighborhood size; OLD20, orthographic Levenshtein distance 20; Root length, root length in letters; log Root Frequency, log-transformed root frequency; log Root Family Size, log-transformed root family size; Root PFMF, percentage of more frequent words in the morphological family of the root; Suffix length, suffix length in letters; log Suffix Frequency, log-transformed suffix frequency; Suffix PFMF, percentage of more frequent words in the morphological family of the suffix; LD latencies. * $p < .05$, ** $p < .01$

Table 4 Standardized β s, R^2 s, and ΔR^2 s for the regression analyses of lexical decision

Step		Root length	log Root Frequency	log Root Family Size	Root PFMF	Suffix length	log Suffix Frequency	Suffix Family Size	Suffix PFMF
Step 1	Length	.039							
	log Frequency	-.386***							
	log Nsize	.053							
	OLD20	.238***							
	R^2	.216							
Step 2	ΔR^2	.216***							
	Imageability	-.076*							
	R^2	.221							
	ΔR^2	.005*							
	All variables but one								
Step 3	Root length	n/a	-.047	-.052	-.057	-.051	-.098**	-.052	-.099**
	log Root Frequency	-.067	n/a	-.078*	-.113**	-.072	-.109*	-.079	-.232***
	log Root Family Size	-.007	-.043	n/a	.007	-.012	.013	-.005	.054
	Root PFMF	-.065	-.090**	-.060	n/a	-.066	-.015	-.054	.065*
	Suffix length	-.052	-.053	-.054	-.056*	n/a	-.052	-.047	-.072*
	log Suffix Frequency	-.337***	-.334***	-.323***	-.306***	-.323***	n/a	-.237***	-.229***
	Suffix Family Size	.115*	.119*	.115*	.107*	.106*	-.117**	n/a	.062
	Suffix PFMF	.297***	.319***	.285***	.251***	.294***	.218***	.272***	n/a
	R^2	.329	.328	.330	.328	.318	.300	.326	.294
	ΔR^2	.108***	.108***	.110***	.107***	.107***	.080***	.106***	.073***
Step 4	Critical variable								
	Variable	-.414	-.073	-.007	-.062	-.053	-.324***	.115*	.287***
	R^2	.330	.330	.330	.320	.320	.330	.330	.330
	ΔR^2	.001	.002	.000	.002	.003	.030***	.004*	.036***

Length, length in syllables; log Frequency, log-transformed frequency; log Nsize, log-transformed orthographic neighborhood size; OLD20, orthographic Levenshtein distance 20; Root length, root length in letters; log Root Frequency, log-transformed root frequency; log Root Family Size, log-transformed root family size; Root PFMF, percentage of more frequent words in the morphological family of the root; Suffix length, suffix length in letters; log Suffix Frequency, log-transformed suffix frequency; Suffix PFMF, percentage of more frequent words in the morphological family of the suffix

Columns refer to the different regression models and their title shows the variables entered in the last step of the models. Critical variable refers to the variable entered in the last step of the regression models. ΔR^2 is the incremental increase in the model R^2 that results from the addition of a predictor or set of predictors in a new step of the model. * $p < .05$ ** $p < .01$ *** $p < .001$

syllables (N-syllables), objective lexical frequency calculated from books (Frequency), imageability (Desrochers & Thompson, 2009), and two measures of orthographic similarity: (1) orthographic neighborhood size (N-size; i.e., the number of words that can be generated by switching one letter: laughter → daughter) and (2) orthographic Levenshtein distance 20 (OLD20) (i.e., the minimum number of insertions, deletions, and substitution required to turn one word into its 20 nearest neighbors) (Yarkoni, Balota, & Yap, 2008). OLD20 has proven to be a good orthographic similarity variable (Cortese & Schock, 2013; Yap & Balota, 2009).

Data analysis

We inspected the data for skewness. Five variables showed skewness values greater than ± 2 (Gravetter & Wallnau, 2014): frequency (3.81), Nsize (2.87), root frequency (23.77), root family size (2.87), and suffix frequency (5.89). We log-transformed these variables, which resolved the skew problems. As can be seen in Table 2, all the variables included in the regression analyses had skewness statistics smaller than ± 2 .

We also inspected the data for multicollinearity. Toward this aim, we ran a regression model entering LD RTs as dependent variable and all the other variables as independent variables. On the basis of this analysis, we calculated the tolerance and variance inflation factor (VIF). The results showed $VIF < 4$ and tolerance $> .2$ for all variables, except for length in letters ($VIF = 5.838$, tolerance = .172). On the basis of the inclusion of another measure of word length in the model (i.e., length in syllables), we decided to exclude length in letters from the model. Thus, the remaining variables showed $VIF < 4$ and tolerance > 0.2 (see Table 2). Furthermore, the majority of the correlation coefficients were less than .75 (Cohen, Cohen, West, & Aiken, 2003; see Table 3). Only one correlation coefficient was greater than .75: the one between log suffix frequency and suffix family, $r(917) = .818$, $p < .01$. This could not be resolved by data transformation. To address this issue, we followed Balota, Cortese, Sergent-Marshall, Spieler, & Yap (2004) and ran two additional regression models. In these models, we excluded one of the two correlated variables to determine whether it influenced the other highly correlated variable entered in the last step of the models. In other words, we ran two additional models, with (1) log suffix frequency entered in the last step and suffix family size excluded, and (2) suffix family size entered in the last step and log suffix frequency excluded. Table 3 shows the correlations between all the variables used as predictors (and the dependent variables) in the LD task.

Following the previous literature (Boukadi, Zouaidi, & Wilson, 2016; Cortese & Schock, 2013; Sánchez-Gutiérrez et al., 2018; Yap & Balota, 2009), we grouped and entered

the variables in the regression models in four separate steps. Step 1 included four lexical variables: length in syllables, log frequency, log NSize, and OLD20. Step 2 included the semantic variable of imageability. Step 3 included all the new morphological variables for the root and suffix, except one ($n = 7$): root length in letters, log root frequency, log root family size, percentage of more frequent words than the target word in its root morphological family, suffix length in letters, log suffix frequency, suffix family size, and percentage of more frequent words than the target word in its suffix morphological family. Step 4, the final step, included each of the eight new morphological variables separately. We thus ran eight different regression models in order to study the specific contribution of each morphological variable above and beyond that of the other variables.

Results

We conducted eight hierarchical regression models with four steps each. In each model, we entered one of the eight new morphological variables at a time in the final step (Step 4). We used LD RTs as the dependent variable. Table 4 shows the results of these analyses. After controlling for the effects of the lexical (Step 1), the semantic (Step 2), and the other seven morphological (Step 3) variables, log suffix frequency, $\beta = -.324$, $p < .001$; suffix family size, $\beta = .115$, $p < .05$; and suffix PFMF, $\beta = .287$, $p < .001$, were all significant predictors of LD latencies. Log suffix frequency had a facilitatory effect on latencies. Suffix family size and suffix PFMF exerted inhibitory effects on latencies. None of the other morphological variables significantly predicted RTs in LD latencies, all $ps > .05$.

In the two additional regression models, run to control for collinearity, the pattern for log suffix frequency remained unchanged, $\beta = -.237$, $p < .001$, with a facilitatory effect on latencies. Suffix family size was still a significant predictor of latencies, but its effect was now facilitatory, $\beta = -.117$, $p < .001$.

Discussion

In this report, we have presented MorphoLex-FR, a sizeable database for derivational morphology in French, based on the 38,840 words of the FLP database (Ferrand et al., 2010) and we also report on analyses with these variables on RTs for 917 morphologically complex nouns from the FLP database. MorphoLex-FR contains twelve new morphological variables, four for each morpheme (i.e., morphological family size, summed token frequency, PFMF and morpheme length for prefixes, roots and suffixes). Studies in French, as in other languages, show disparate results in terms of the effects of morphological variables in visual word processing (Amenta

& Crepaldi, 2012; Feldman & Milin, 2018). We believe that one of the main reasons for these inconsistent results resides in the fact that calculations for these morphological variables are not comparable across studies (Sánchez-Gutiérrez et al., 2018). In that regard, the main contribution of the present study is the inclusion of calculations for the most studied morphological variables in visual word processing literature. We based these calculations on a sizeable corpus of French words thus assuring their representativeness and reliability. The public online availability of MorphoLex-FR will render future studies in French comparable, offering an opportunity for researchers to investigate the effects of these new morphological variables in a wide range of tasks and a variety of experimental designs in French. Moreover, the ways in which the calculations for these variables were made for MorphoLex-FR (French) and MorphoLex (English; Sánchez-Gutiérrez et al., 2018) are identical. This will facilitate the comparison between French and English future studies and theorizing on morphological processing.

Indicating the validity of the database, we found that the database roughly reproduced the distribution of morphologically simple and complex words in French, with 41% of the words being monomorphemic, and 59% of the words being morphologically complex (Rey-Debove, 1984). Among the morphologically complex words in MorphoLex-FR, the most frequent type of morphologically complex words were those with one suffix (e.g., *travailleur* [worker], with its root *travail* [work] and its suffix *-eur* [-er]). This PRS signature represents almost as many tokens as the morphologically simple words in the FLP database. The other PRS signatures are fairly less frequent and represent altogether 22% of the words. The roots in morphologically complex words exhibited larger family sizes and larger summed token frequency than simple words. In addition, and as expected, the roots in morphologically complex words tended to have more frequent members in their morphological families than simple words. Also, simple and complex words showed comparable root lengths (in number of letters), though slightly longer for simple words. In the realm of morphologically complex words, the comparison between prefixes and suffixes indicated that suffixed words tended (1) to be longer than prefixed words in terms of their number of letters, (2) to belong to larger morphological families, and (3) to be more frequent. Altogether, this suggests that suffixes might be more salient than prefixes, a pattern reflected in studies to date in, for instance, Italian and Dutch (Burani & Thornton, 2003; Kuperman, Bertram, & Baayen, 2010; Laudanna & Burani, 1995). In a previous study in English, we showed that words with more salient suffixes (i.e., longer, more frequent and from larger family sizes) were processed faster than words with less salient suffixes (Sánchez-Gutiérrez et al., 2018). The results of the experimental LD study that we conducted here strongly suggest that this is also the case for French.

To further validate this database, we studied the influence of the new morphological French variables on LD latencies obtained from the FLP. We entered these new morphological variables and control variables as predictors of LD latencies for 917 suffixed nouns in a series of hierarchical regression models. Results indicate that morphological variables related to the suffix (but not to the root) explained LD RTs. Specifically, the higher the frequency of a suffix and the larger the family size of a suffix, the shorter the latencies. In accordance with our results, previous studies have also found facilitatory effects of suffix frequency (Baayen et al., 2007; Burani & Thornton, 2003; Sánchez-Gutiérrez et al., 2018) and suffix family size (Bertram, Baayen, & Schreuder, 2000; Lázaro & Sainz, 2012; Sánchez-Gutiérrez et al., 2018). It is noteworthy that in our study in French, as in our parallel analyses in English (Sánchez-Gutiérrez et al., 2018), suffix frequency and suffix family size were highly associated. This means that suffixes from larger families also tend to be more frequent, and vice versa (Burani & Thornton, 2003). In the present study, the absence of suffix frequency in the model actually changed the direction of the influence of suffix family size on visual word recognition (from inhibitory to facilitatory). Also, we found an inhibitory effect of the percentage of more frequent words in the morphological family of the suffix. That is to say, the higher the percentage of more frequent words in the morphological family of the suffix, the longer the latencies. Our results are in line with previous studies conducted in French (Colé et al., 1989; Meunier & Segui, 1999) and English (Sánchez-Gutiérrez et al., 2018). In other words, it seems that when a word belongs to a large morphological affixal family and that the affix is frequent, this facilitates word recognition. Conversely, if a word has many suffix family members that are more frequent words, competition increases and recognition is thus slowed.

Interestingly, and unlike previous studies, we failed to find an influence of root morphological variables on speed of word recognition. For instance, a facilitatory effect of root frequency in word recognition has been consistently reported in French (Colé et al., 1989) and other languages (Burani & Thornton, 2003; Caramazza, Laudanna, & Romani, 1988; Colé et al., 1989; Luke & Christianson, 2011; Taft & Forster, 1976). Several factors can account for this difference between our study and those in previous literature. First, most of the previous studies that addressed root morphological variables were factorial. Controlling for other morphological variables in our regression design we were able to evaluate the influence of these morphological variables all together. In our paradigm, in which both root and suffix morphological variables were simultaneously present, only suffix morphology affected word recognition. Schreuder and Baayen (1997) also failed to find that root frequency affected word recognition in Dutch monomorphemic words. Another possible explanation may come from cross-linguistic

differences, specifically from the characteristics of French roots as compared to those in English. This is worth considering, given that parallel analyses in English showed that both root frequency and root family size benefited word recognition in English, even in the presence of suffix morphological variables (Sánchez-Gutiérrez et al., 2018). Contrary to French, many English roots are embedded in derived words as *obsession*; for instance, for the noun *obsession*, the English root is *obsess*, whereas the French root is *obséd-*. This may disfavor root morphology processing to a larger extent in French than in other languages. This substantive influence of suffix morphology both in this study of French and in our prior work in English (Sánchez-Gutiérrez et al., 2018) is perhaps counterintuitive, but points to the need to further study its influence. Only the study of both root and suffix morphology simultaneously will allow us to have an accurate and complete depiction of morphological processing (see, e.g., Kuperman, Bertram, & Baayen, 2008, for trimorphemic Finnish compounds). Indeed, in a similar study as the one we present here, we studied visual word recognition of English words that contained both a prefix and suffix by means of a written LD task (Wilson, Sánchez-Gutiérrez, Mailhot, & Deacon, 2019). After controlling for the effects of lexical and semantic variables, root cumulative frequency and prefix productivity exerted a facilitatory effect. The percentage of more frequent words than the target in the families of the prefix and the suffix had an inhibitory effect. Our results support the contribution of root frequency. They also extend to prefix morphological variables the findings of the present study on the influence of suffix morphology.

Among the limitations of the database is the fact that we applied a criterion of semantic transparency to segment words. Consequently, words that have the appearance of being morphologically complex but that are monomorphemic—that is, pseudo-derived words—were segmented as being monomorphemic. For instance, a morphologically complex word as *tablette* [little table] was segmented as (*table*)>*ette*>, thus as having one root (i.e., *table*) and one suffix (i.e., *-ette*). Conversely, a pseudo-derived word as *baguette* [baguette; in French, *bague* means “ring,” so *baguette* could be literally interpreted as “little ring”] was segmented as being monomorphemic (*baguette*). Experimental evidence has shown facilitatory priming effects for words that share an apparent morphological form (i.e., pseudo-derived words) but that are not semantically related, such as *baguette*–*bague* [baguette–ring] or the English equivalent *corner*–*corn* (Emmorey, 1989; Forster & Azuma, 2000; Pastizzo & Feldman, 2004; Rastle et al., 2004). This pattern of results has also been found in French. Giraudo and Voga (2016) showed that French words sharing the same bound-stem primed each other. Quémart and Casalis (2015) found that the orthographic overlap between word pairs exerted an effect in normally developing French-speaking children, whereas the semantic properties of the

morphemes primed written word processing in dyslexic children. Taken together, these results suggest that the recognition of a word can occur by means of a segmentation based solely on the morpho-orthographic properties of a word (Rastle & Davis, 2008; Rastle et al., 2004). Consequently, this has been taken as evidence that semantic transparency is not critical, at least for priming paradigms (Forster & Azuma, 2000). The evidence for this morpho-orthographic priming in the absence of semantic effects comes almost exclusively from studies using the masked priming paradigm. Therefore, the present database might be of limited usefulness for researchers interested in the study of the orthographic overlap between words but without semantic relatedness.

In sum, MorphoLex-FR takes a first step in addressing the lack of shared sizeable databases necessary for reproducible and comparable research on morphological processing in French. The procedure we used for morphological segmentation and variable computation is consistent and explicit. This renders MorphoLex-FR a suitable instrument for large-scale studies that will contribute to our understanding of how morphologically complex words are processed in French.

Acknowledgements This research was supported by an Insight Development Grant awarded to M.A.W., S.H.D., and C.S.G. by the Social Sciences and Humanities Research Council (CRSH) of Canada, grant number 430-2015-00699.

References

- Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41–61. <https://doi.org/10.1006/jmla.1998.2607>
- Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, 3, 232:1–12. <https://doi.org/10.3389/fpsyg.2012.00232>
- Apothéloz, D. (2002). *La construction du lexique français: principes de morphologie dérivationnelle*. Paris, France: Ophrys.
- Auclair-Ouellet, N., Fossard, M., Houde, M., Laforce, R., & Macoir, J. (2016). Production of morphologically derived words in the semantic variant of primary progressive aphasia: preserved decomposition and composition but impaired validation. *Neurocase*, 22, 170–178. <https://doi.org/10.1080/13554794.2015.1081391>
- Auclair-Ouellet, N., Fossard, M., Laforce, R., Bier, N., & Macoir, J. (2017). Conception or *conceivation? The processing of derivational morphology in semantic dementia. *Aphasiology*, 31, 166–188. <https://doi.org/10.1080/02687038.2016.1168918>
- Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313. <https://doi.org/10.1016/j.jml.2006.03.008>
- Baayen, R. H., Wurm, L. H., & Aycok, J. (2007). Lexical dynamics for low-frequency complex words. A regression study across tasks and modalities. *Mental Lexicon*, 2, 419–463. <https://doi.org/10.1075/ml.2.3.06baa>
- Balling, L., & Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive*

- Processes*, 23, 1159–1190. <https://doi.org/10.1080/01690960802201010>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133(2), 283–316. <https://doi.org/10.1037/0096-3445.133.2.283>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459. <https://doi.org/10.3758/BF03193014>
- Bauer, L., & Nation, P. (1993). Word Families. *International Journal of Lexicography*, 6, 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language*, 35, 801–820. <https://doi.org/10.1006/jmla.1996.0041>
- Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, 42, 390–405. <https://doi.org/10.1006/jmla.1999.2681>
- Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review*, 22, 1054–1061. <https://doi.org/10.3758/s13423-014-0752-9>
- Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, 20, 220–230. <https://doi.org/10.1080/10888438.2016.1140769>
- Beyersmann, E., Iakimova, G., Ziegler, J. C., & Colé, P. (2014). Semantic processing during morphological priming: An ERP study. *Brain Research*, 1579, 45–55. <https://doi.org/10.1016/j.brainres.2014.07.010>
- 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>
- Booij, G. E. (2005). Compounding and derivation: Evidence for construction morphology. In W. U. Dressler, F. Rainer, D. Kastovsky, & O. Pfeiffer (Eds.), *Morphology and its demarcations* (pp. 109–132). Amsterdam, The Netherlands: Benjamins.
- Booij, G. E. (2006). Inflection and derivation. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd, pp. 654–661). Oxford, UK: Elsevier.
- Boukadi, M., Zouaidi, C., & Wilson, M. A. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods*, 48, 585–599. <https://doi.org/10.3758/s13428-015-0602-3>
- Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 157–208). Berlin, Germany: Mouton de Gruyter.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332. <https://doi.org/10.1016/0010-027790017-0>
- Carstairs-McCarthy, A. (2006). Affixation. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd, pp. 83–88). Oxford, UK: Elsevier.
- Cartoni, B., & Lefer, M. A. (2010). *The MuLexFoR database: Representing word-formation processes in a multilingual lexicographic environment*. Paper presented at the International Conference on Language Resources and Evaluation, Valletta, Malta.
- Cavalli, E., Colé, P., Pattamadilok, C., Badier, J. M., Zielinski, C., Chanoine, V., & Ziegler, J. C. (2017). Spatiotemporal reorganization of the reading network in adult dyslexia. *Cortex*, 92, 204–221. <https://doi.org/10.1016/j.cortex.2017.04.012>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd). New York, NY: Routledge.
- Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, 28, 1–13. [https://doi.org/10.1016/0749-596X\(89\)90025-9](https://doi.org/10.1016/0749-596X(89)90025-9)
- Corbin, D. (2004). *La formation des mots: Horizons actuels*. Villeneuve d'Ascq, France: Presses universitaires du Septentrion.
- Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, 66, 946–972. <https://doi.org/10.1080/17470218.2012.722660>
- D'Alessio, M. J., Jaichenco, V., & Wilson, M. A. (2018). The role of morphology in word naming in Spanish-speaking children. *Applied Psycholinguistics*, 39, 1065–1093. <https://doi.org/10.1017/s0142716418000127>
- Deacon, S. H., & Francis, K. A. (2017). How children become sensitive to the morphological structure of the words that they read. *Frontiers in Psychology*, 8, 1469. <https://doi.org/10.3389/fpsyg.2017.01469>
- Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, 18, 432–451.
- Desrochers, A., & Thompson, G. L. (2009). Subjective frequency and imageability ratings for 3,600 French nouns. *Behavior Research Methods*, 41, 546–557. <https://doi.org/10.3758/BRM.41.2.546>
- Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, 20, 75–114. <https://doi.org/10.1080/01690960444000197>
- Emmorey, K. D. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4, 73–92. <https://doi.org/10.1080/01690968908406358>
- Feldman, L. B., & Milin, P. (2018). Psycholinguistic studies of word morphology and their implications for models of the mental lexicon and lexical processing. In R. Berthiaume, D. Daigle, & A. Desrochers (Eds.), *Morphological processing and literacy development: Current issues and research* (pp. 16–47). New York, NY: Routledge.
- Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, 9, 111. <https://doi.org/10.3389/fnhum.2015.00111>
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- Ford, M. A., Davis, M. H., & Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, 63, 117–130. <https://doi.org/10.1016/j.jml.2009.01.003>
- Forster, K. I., & Azuma, T. (2000). Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, 15, 539–561. <https://doi.org/10.1080/01690960050119698>
- Gala, N., & Rey, V. (2008, June). *Polymots: Une base de données de constructions dérivationnelles en français à partir de radicaux phonologiques*. Paper presented at the Conférence sur le Traitement Automatique des Langues Naturelles—TALN, Senlis, France.
- Giraud, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming.

- Language and Cognitive Processes*, 15, 421–444. <https://doi.org/10.1080/01690960050119652>
- Giraud, H., & Voga, M. (2016). Words matter more than morphemes: An investigation of masked priming effects with complex words and non-words. *Italian Journal of Linguistics*, 28, 49–78.
- Gravetter, F. J., & Wallnau, L. B. (2014). *Essentials of statistics for the behavioral sciences* (8th). Belmont, CA: Wadsworth.
- Hockett, C. F. (1958). *A course in modern linguistics*. New York, NY: Macmillan.
- Kuperman, V., Bertram, R., & Baayen, H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, 62, 83–97. <https://doi.org/10.1016/j.jml.2009.10.001>
- Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23, 1089–1132. <https://doi.org/10.1080/01690960802193688>
- Lachaud, C. M. (2007). CHACQFAM: Une base de données renseignant l'âge d'acquisition estimé et la familiarité pour 1225 mots monosyllabiques et bisyllabiques du français. *Année Psychologique*, 107, 39–63.
- Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing: Cross-linguistic perspectives* (pp. 345–364). Hillsdale, NJ: Erlbaum.
- Lázaro, M., & Sainz, J. S. (2012). The effect of family size on Spanish simple and complex words. *Journal of Psycholinguistic Research*, 41, 181–193. <https://doi.org/10.1007/s10936-011-9186-y>
- Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, 35, 121–146. <https://doi.org/10.1007/s10936-005-9008-1>
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156–166. <https://doi.org/10.3758/bf03195560>
- Longtin, C. M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, 53, 26–41. <https://doi.org/10.1016/j.jml.2005.02.008>
- Longtin, C. M., Segui, J., & Halle, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18, 313–334. <https://doi.org/10.1080/01690960244000036>
- Luke, S. G., & Christianson, K. (2011). Stem and whole-word frequency effects in the processing of inflected verbs in and out of a sentence context. *Language and Cognitive Processes*, 26, 1173–1192. <https://doi.org/10.1080/01690965.2010.510359>
- Marcolini, S., Traficante, D., Zoccolotti, P., & Burani, C. (2011). Word frequency modulates morpheme-based reading in poor and skilled Italian readers. *Applied Psycholinguistics*, 32, 513–532. <https://doi.org/10.1017/S0142716411000191>
- Marslen-Wilson, W. (2006). Morphology and language processing. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd, pp. 295–300). Oxford, UK: Elsevier.
- Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3–33. <https://doi.org/10.1037/0033-295X.101.1.3>
- McCormick, S. F., Brysbaert, M., & Rastle, K. (2009). Is morphological decomposition limited to low-frequency words? *Quarterly Journal of Experimental Psychology*, 62, 1706–1715. <https://doi.org/10.1080/17470210902849991>
- Meunier, F., & Longtin, C. M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56, 457–471. <https://doi.org/10.1016/j.jml.2006.11.005>
- Meunier, F., & Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, 41, 327–344. <https://doi.org/10.1006/jmla.1999.2642>
- Morris, J., Porter, J. H., Grainger, J., & Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language and Cognitive Processes*, 26, 558–599. <https://doi.org/10.1080/01690965.2010.495482>
- Moscato del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30, 1271–1278. <https://doi.org/10.1037/0278-7393.30.6.1271>
- Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, 47, 3–12. <https://doi.org/10.1177/0022219413509967>
- Namer, F. (2003). Automatiser l'analyse morpho-sémantique non affixale: Le système DérIF [To automatize non-affixal morpho-syntactic analysis: The DérIF system]. In N. Hathout, M. Roché, & N. Serna (Eds.), *Cahiers de grammaire* (pp. 31–48). Toulouse, France: ERSS.
- Namer, F. (2013). A rule-based morphosemantic analyzer for French for a fine-grained semantic annotation of texts. *Communications in Computer and Information Science*, 380, 92–114. https://doi.org/10.1007/978-3-642-40486-3_6
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28, 661–677. <https://doi.org/10.1017/S014271640707035X>
- 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>
- New, B., Pallier, C., & Ferrand, L. (2005). Lexique 3 [Database]. Retrieved from www.lexique.org
- Oganyan, M., Wright, R., & Herschensohn, J. (2019). The role of the root in auditory word recognition of Hebrew. *Cortex*, 116, 286–293. <https://doi.org/10.1016/j.cortex.2018.06.010>
- Paice, C. D. (2006). Stemming. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd, pp. 149–150). Oxford, UK: Elsevier.
- Pastizzo, M. J., & Feldman, L. B. (2004). Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language*, 90, 31–39. [https://doi.org/10.1016/S0093-934X\(03\)00417-6](https://doi.org/10.1016/S0093-934X(03)00417-6)
- Peereman, R., Sprenger-Charolles, L., & Messaoud-Galusi, S. (2013). The contribution of morphology to the consistency of spelling-to-sound relations: A quantitative analysis based on French elementary school readers. *Année Psychologique*, 113, 3–33. <https://doi.org/10.4074/s0003503313001012>
- Quémart, P., & Casalis, S. (2015). Visual processing of derivational morphology in children with developmental dyslexia: Insights from masked priming. 36, 345–376. <https://doi.org/10.1017/S014271641300026X>
- Quémart, P., Casalis, S., & Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology*, 109, 478–496. <https://doi.org/10.1016/j.jecp.2011.02.008>
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23, 942–971. <https://doi.org/10.1080/01690960802069730>
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11, 1090–1098. <https://doi.org/10.3758/BF03196742>
- Rey-Debove, J. (1984). Le domaine de la morphologie lexicale. *Cahiers de Lexicologie*, 45, 3–19.
- Romary, L., Salmon-Alt, S., & Francopoulo, G. (2004). *Standards going concrete: From LMF to Morphalou*. Paper presented at the 20th

- International Conference on Computational Linguistics—COLING 2004, Geneva, Switzerland.
- Sánchez-Gutiérrez, C. (2013). Contribución de la conciencia morfológica a la mejora de la ortografía: Un estudio evolutivo con niños de tercero a sexto de primaria. *Sintagma*, 25, 33–46.
- Sánchez-Gutiérrez, C., & Hernández-Muñoz, N. (2018). Development of derivational morphological awareness in anglophone learners of Spanish: A relational knowledge study. *Foreign Language Annals*, 51, 369–388. <https://doi.org/10.1111/flan.12344>
- Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. E., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior Research Methods*, 50, 1568–1580. <https://doi.org/10.3758/s13428-017-0981-8>
- Schreuder, R., & Baayen, R. H. (1997). How simplex complex words can be. *Journal of Memory and Language*, 37, 118–139. <https://doi.org/10.1006/jmla.1997.2510>
- Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4, 353–361. <https://doi.org/10.1016/S1364-661301515-1>
- Sparks, E., & Deacon, S. H. (2015). Morphological awareness and vocabulary acquisition: A longitudinal examination of their relationship in English-speaking children. *Applied Psycholinguistics*, 36, 299–321. <https://doi.org/10.1017/S0142716413000246>
- Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, 14, 17–26. <https://doi.org/10.3758/BF03209225>
- Suárez-Coalla, P., Martínez-García, C., & Cuetos, F. (2017). Morpheme-based reading and writing in Spanish children with dyslexia, 8, 1952. <https://doi.org/10.3389/fpsyg.2017.01952>
- Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *Mental Lexicon*, 1, 183–199. <https://doi.org/10.1075/ml.1.2.02taf>
- Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15, 607–620. <https://doi.org/10.1016/0022-537190054-2>
- Wilson, M. A., Sánchez-Gutiérrez, C. H., Mailhot, H., & Deacon, S. H. (2019). The importance of being earnest in earnestness: The influence of root morphology in complex noun processing. In S. Sulpizio, L. Barca, S. Primativo, & L. S. Arduino (Eds.), *Word recognition, morphology and lexical reading: Essays in honour of Cristina Burani* (pp. 134–143). London, UK: College Publications.
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60, 502–529. <https://doi.org/10.1016/j.jml.2009.02.001>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. <https://doi.org/10.3758/pbr.15.5.971>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.