Masked constituent priming of English compounds in native and nonnative speakers

Article

Accepted Version

Masked constituent priming of English compounds in native and nonnative speakers

Jorge González Alonso[1], Silvia Baquero Castellanos[2] & Oliver Müller[3]

1. University of Reading, UK, 2. Universidad Nacional, Colombia, 3. Universidad del Rosario, Colombia

The present research explores the degree of morphological structure of compound words in the native and nonnative lexicons, and provides additional data on the access to these representations. Native and nonnative speakers (L1 Spanish) of English were tested using a lexical decision task with masked priming of the compound's constituents in isolation, including two orthographic conditions to control for a potential orthographic locus of effects. Both groups displayed reliable priming effects, unmediated by semantics, for the morphological but not the orthographic conditions as compared to an unrelated baseline. Results contribute further evidence of morphological structure in the lexicon of native speakers, and suggest that lexical representation and access in a second language are qualitatively comparable at relatively advanced levels of proficiency.

Keywords**:** lexical representation; non-native speakers; compounds; morphological processing; masked priming.

Introduction

Among the considerable issues in visual word recognition research, the role of morphology is hotly debated. While many agree that, at least in some cases, native visual word processing involves morphological decomposition for complex words (e.g., Marslen-Wilson, 2007), there is no consensus as to exactly how, when, and under what conditions this happens. As with any other aspect of language processing, access is crucially interesting for psycholinguists in that it is inextricably bound to representation. In fact, the existence of—or even the need for—a morphologically structured lexicon has been the subject of much discussion (see, e.g., Feldman, O'Connor, & Moscoso del Prado Martín, 2009). Focused on the visual processing of English compound words in

isolation, the data reported in this article provide supporting evidence for models assuming some degree of morphological structure in the lexicon of native speakers, and suggests that lexical representations in relatively proficient second language users are equally complex.

The role constituent morphemes play in the access to, and representation of, complex words has been explored systematically for the last 40 years, from the seminal studies of Taft and Forster (Taft & Forster, 1975, 1976) to the proliferation of research contrasting behavioural, electrophysiological, neuroimaging and clinical data (e.g., Bozic, Marslen-Wilson, Stamatakis, Davis, & Tyler, 2007; Fiorentino, Naito-billen, Bost, & Fund-Reznicek, 2014; MacGregor & Shtyrov, 2013; Marelli, Zonca, Contardi, & Luzzatti, 2014; Marslen-Wilson & Tyler, 2007, among others). Current positions range from models advocating exclusively morpheme mediated access (Stockall & Marantz, 2006; Taft & Nillsen, 2013), to dual/multiple-route accounts highlighting the competition (and/or cooperation) between whole-form and morpheme-based access (Hyönä, 2012), to proposals questioning the role of morphology in favour of a distributed-connectionist implementation of unmediated access to constituent and whole-word meanings through orthographic and/or phonological cues (e.g., Devlin, Jamison, Matthews, & Gonnerman, 2004; Kuperman, 2013; Seidenberg & Gonnerman, 2000).

The masked priming experimental paradigm (Forster & Davis, 1984), often combined with a lexical decision task, has been widely used by researchers ascribing to most, if not all, of the above accounts (e.g., Feldman, Kostić, Basnight-Brown, Durđević, & Pastizzo, 2010; Rastle & Davis, 2008). In this technique, the target word is preceded by two items: the first is a forward mask, consisting of a string of symbols (e.g., '#') matching the prime in length, and displayed for approximately 500

milliseconds; immediately afterwards, the prime is presented for a very short lapse, which varies widely (typically 34-60 ms; Rastle & Davis, 2008) but rarely exceeds 80 ms (Neubauer & Clahsen, 2009), in order to prevent conscious processing. Through an appropriate manipulation of the prime-target SOA, the paradigm can provide insight into the time course of access to different types of linguistic information—i.e., orthographic, morphological, semantic, etc. Data from masked priming experiments has thus been used to weigh in on the debate as to whether morphological effects originate in lexical structure or are the reflex of a systematic relationship between formal (orthographic/phonological) and semantic overlap. Results in this paradigm often contrast with what is observed with overt priming. In a series of cross-modal priming experiments, Marslen-Wilson, Tyler, Waksler and Older (1994) found that morphological effects obtained even with little formal overlap (e.g., *elusive – ELUDE*), but never in the absence of a semantic relationship (e.g., *casualty – CASUAL*). When the overlap dimension was purely formal, no priming effects were observed (e.g., *tinsel – TIN*). In contrast, some studies have reported orthographic priming effects at short stimulus onset asynchronies (SOAs) under masked priming conditions (e.g., Chateau, Knudsen, & Jared, 2002; Forster & Azuma, 2000; Grainger, Granier, Farioli, Van Assche, & van Heuven, 2006). In three masked priming experiments manipulating prime duration, Forster and Azuma (2000) compared priming effects between a) prefixed words sharing a bound stem (*submit – PERMIT*), b) a prefixed word and its free stem (*fold – UNFOLD*), and c) unrelated words sharing their (non-morphological) final graphemes (*shallow – FOLLOW*). Priming effects were comparable for conditions (a) and (b) with a prime duration of 50 ms. Crucially, with this SOA a priming effect emerged for condition (c) as well. In order to determine whether the priming effect in (a) was orthographic or morphological in nature, the authors increased the difficulty of

the task (by making non-words more word-like, and thus harder to discriminate) and raised prime duration to 68 ms. With these modifications, priming effects disappeared for the orthographic condition (c), but remained for the bound and free morphological conditions, (a) and (b). Other studies employing the same paradigm have reported inhibitory effects of orthographically related primes (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Frisson, Bélanger, & Rayner, 2014). Considered together, these results suggest that, at an early stage and at least under certain conditions, both morphological and orthographic priming effects may obtain, and thus be confounded, between words. Overall, however, the emergence of orthographic priming effects—and whether these are inhibitory or facilitatory—seems to be conditioned by a large number of factors, which include prime duration and visibility, lexicality and neighbourhood density of the prime, nature and relative position of the coincidental letters within the string, and nature of the experimental task, among others (see, e.g., Carreiras, Armstrong, Perea, & Frost, 2014; Frisson, Bélanger, & Rayner, 2014; Frost, Kugler, Deutsch, & Forster, 2005; Grainger, Kiyonaga, & Holcomb, 2006; Ktori, Midgley, Holcomb, & Grainger, 2015; Lee, Rayner, & Pollatsek, 1999; Robert & Mathey, 2012; van Heuven, Dijkstra, Grainger, & Schriefers, 2001). Despite this seeming dissociation of morphological and orthographic effects in the literature, an orthographic prime condition was included in the present study—together with constituent to compound transparency ratings—to help determine if any potential morphological effects actually emerge from a combination of formal and semantic overlap.

Among the three word-formation processes traditionally considered in the literature, namely, inflection, derivation and compounding, the latter has received increased attention in recent years (see, e.g., Semenza & Luzzatti, 2014, for review). Compounding differs from the other two processes in nontrivial ways. First, unlike

inflection (and similar to derivation), compounding is assumed to create new lexical entries, which may acquire idiosyncratic meanings with a variable distance from their predicted, compositional meaning. Second, unlike both derivation and inflection, compounding entails the combination of two or more free morphemes. These characteristics make compounding particularly useful to address perennial questions in the visual word processing literature, for at least two reasons. The first is that more than one lexical meaning is activated, which means that potentially very different semantic features will be activated independently. This provides insight into the time-course of semantic effects in morphological processing. The second reason is that, being a (mostly unmediated) combination of free morphemes, most languages lack functional markers associated to compounding. This makes data from compounding studies showing rapid morphological decomposition particularly compelling, since parsing mechanisms cannot resort to affix detection on the basis of frequency and regularity (Fiorentino et al., 2014).

Constituent priming effects have been reliably found in both overt (e.g., Libben, Gibson, Yoon, & Sandra, 2003; Sandra, 1990; Zwitserlood, 1994) and masked priming experiments (e.g., Duñabeitia, Laka, Perea, & Carreiras, 2009; Fiorentino & Fund-Reznicek, 2009; Shoolman & Andrews, 2003). Libben et al. (2003) conducted a series of experiments employing both overt constituent priming and standard (i.e., unprimed) lexical decision tasks, in order to explore the processing of transparent (e.g., *snowball*), partially transparent (e.g., *strawberry*) and opaque (e.g., *bellhop*) compounds in English. They found that priming effects obtained from both constituents for all compound types, and that response times (RTs) were strongly affected by the semantic transparency of the head constituent, with opaque heads yielding longer latencies. Focusing on early effects, Duñabeitia et al. (2009) conducted a series of lexical decision experiments with

masked priming using Basque compounds. In their study, target compounds could be preceded by other compounds sharing the first or second constituent, in either the same (Experiment 1) or the opposite (Experiment 2) position (e.g., *snowman – SNOWBALL* or *snowman – POSTMAN*, Experiment 1; *football – BALLPOINT* or *football – BAREFOOT*, Experiment 2). Both tasks showed reliable constituent priming effects for both constituents, irrespective of position, head directionality or transparency[1].

Non-compositional views of the lexicon have been significantly challenged by studies on compound processing focusing on representation. Shoolman and Andrews (2003) utilised the masked priming paradigm to investigate constituent priming effects on both transparent (e.g., *snowball*) and opaque (e.g., *jailbird*) compound words as compared to monomorphemic, yet seemingly complex words (e.g., *hammock*), and various kinds of pseudocompounds—with an orthogonal manipulation of lexicality in the first and second constituents—to examine nonword context effects. Results showed reliable masked priming effects for both constituents in compound words, irrespective of context and transparency, as well as in pseudo-complex monomorphemic words (i.e., both *ham* and *mock* significantly enhanced the classification of *hammock*). Shoolman and Andrews's (2003) results mirror some of the findings in the masked priming literature on derivational word-formation (i.e., transparent, opaque and pseudo-complex items all show constituency effects, with the latter displaying smaller priming effects; see also Diependaele, Duñabeitia, Morris, & Keuleers, 2011), and speak to both decomposition and representation. The advantage of compounds over monomorphemes was constant irrespective of semantic transparency and in spite of the higher overall frequency of the latter category (see Fiorentino & Poeppel, 2007; Ji, Gagné, & Spalding, 2011, for similar results), which indicates that activation of constituents is associated with the activation of whole-word forms without mediation from semantics.

Research on nonnative language learners has not yet reached an agreement as to how much morphological information is present in the non-native lexicon, and particularly how/whether this information might be accessed in real time. Evidence has been provided in support of two opposing views, namely, that native and nonnative processing of morphology may differ quantitatively (in processing speed and capacity) but not qualitatively (e.g., Diependaele et al., 2011; Duñabeitia, Dimitropoulou, Morris, & Diependaele, 2013), and that native and nonnative speakers do employ substantially different mechanisms (e.g., Clahsen, Balkhair, Schutter, & Cunnings, 2013).

Feldman, Kostić, Basnight-Brown, Durđević and Pastizzo (2010) compared the performance of native speakers and two groups of nonnative speakers (high and low proficiency) in masked and cross-modal priming experiments with regularly and irregularly inflected verb forms as primes, and their infinitives as targets. Both native and nonnative speakers showed priming effects in all conditions. This pattern seemed to be driven by their high proficiency subjects, whose results were much more consistent than those of the low proficiency sub-group. The findings were interpreted as evidence for a non-decompositional, single-mechanism theory of morphological processing for natives and nonnatives, irrespective of proficiency. Differences between L1 and low-proficiency L2 participants are attributed to deficits in the nonnative speakers' associations of shared form and meaning between different inflections due to insufficient or inappropriate input, rather than to a fundamental difference in the way native and nonnative speakers process these words.

Similarly, the results of the nonnative participants in the Diependaele et al. (2011) study support models postulating native and nonnative speakers' use of the same underlying mechanisms, irrespective of what these mechanisms are assumed to be. An inspection of response latencies in a lexical decision task with masked priming of root

forms by derivations revealed that two groups of nonnative speakers which differed in their L1 but had similar proficiency in English (Spanish and Dutch natives, lower-advanced) showed the same patterns of priming as native speakers. In particular, both groups had larger priming effects for transparent (*walker – WALK*) than opaque (*corner – CORN*) derivations as compared to an unrelated baseline. Notably, though, nonnative speakers showed also a small but significant effect of non-morphological form priming (*freeze – FREE*), which was absent for native speakers (as is typically the case, see, e.g., Rastle & Davis, 2008). Diependaele et al. (2011) propose that the effect may stem from a delay in the orthographic level due to slower processing abilities in the L2 more generally, which would in turn delay the activation of the lexical and morpho-semantic features that block the orthographic effect in native speakers.

Duñabeitia et al. (2013) further investigated the role of proficiency in bilingual processing by means of a masked cross-language morphological priming task, in which a given stem (e.g., *pain*) could be primed by a derived word in the target language (e.g., *painful*) or a word derived from the stem's translation in the non-target language (e.g., *doloroso*, lit. *painful*, whose morphemic structure is {dolor}, lit. *pain*, + {-oso}, an adjectival suffix). Two groups were examined: balanced Basque-Spanish bilinguals (early acquirers of both languages) and unbalanced Spanish-English bilinguals (lower-advanced L2 learners), performing the task in Spanish and English, respectively. Two findings stand out as particularly relevant: 1) both groups showed reliable priming from affixed words to stems in within-language trials (e.g., *doloroso – DOLOR*; *painful – PAIN*), and 2) none of the two groups showed effects of translation priming before or after decomposition, suggesting that early and automatic morpho-orthographic analysis happens for native speakers, balanced and unbalanced bilinguals alike. While these results, like the ones Diependaele et al. (2011), do not rule out the possibility of a

qualitatively different behaviour in low-proficiency levels, they could be interpreted as evidence that, beyond a given stage in L2 development, native-like processing (in qualitative terms) is not a function of proficiency or age of acquisition (AoA).

Conversely, differences between native and nonnative populations have been often reported (see Clahsen et al., 2013, for review). Silva and Clahsen (2008) conducted four masked priming experiments on regular inflection and {-ness} and {-ity} adjectival derivations, with (lower-) advanced L2 learners of three different mother tongues (Chinese, Japanese and German) and native English controls as participants. Results show that nonnative speakers fail to reveal priming effects with inflectional forms (experiments 1 and 2), although Silva and Clahsen do report reduced priming of stems by derived forms in the L2 groups (experiments 3 and 4). In a later experiment, Clahsen et al. (2013, experiment 1) employed an alternative design to determine if the lack of stem priming shown by L2 speakers may be due to slower processing. The new experiment included an inter-stimulus interval of 500 ms post-prime display, thus allowing a longer time for the prime to be processed. The new group of nonnative speakers, who had Arabic as their L1 and were equally lower-advanced in English, failed to show stem priming effects with regular inflections in either version of the experiment. Similarly, Neubauer and Clahsen (2009) found comparable masked priming effects for L1 and L2 speakers of German (L1 Polish) with irregular inflections only. Moreover, the authors found surface frequency effects of regularly inflected participles for the nonnative speaker group only. Taken together, Clahsen and colleagues interpret these findings as evidence that inflection and derivation are essentially different processes (cf. discussion in Clahsen, Sonnenstuhl, & Blevins, 2003), and that nonnative speakers, who are unable to use morphological information in real time in the same way as native speakers, store regularly inflected words in whole form yet have

morphologically structured representations for derived words—which they employ less efficiently than natives.

In an eye-tracking study, Cunnings and Clahsen (2008) demonstrated that native speakers displayed earlier sensitivity to a constraint that selects singular over plural forms to appear inside compounds (the so-called Category Constraint), with the effects of a morphological constraint (a Structural Constraint) that further allows the incorporation of only irregular plurals showing up in later eye-movement measures. Clahsen et al. (2013) replicate the experiments in Cunnings and Clahsen (2008) with second language learners. In an offline (grammaticality judgement) and an online (eye-tracking) task, 25 nonnative speakers of English (L1 Dutch) were presented with sentences containing derived words (adjectival {-less} and adverbial {-wise} derivations) with base nouns that were either singulars (e.g., *flealess*), regular plurals (e.g., *fleasless*) or irregular plurals (e.g., *liceless*). Despite showing native-like preferences in the off-line task (i.e., banning regular plurals and preferring singular over irregular plural forms), there were some differences between the eye-movements of the nonnative speakers in Clahsen et al. (2013) and those of the native speakers in Cunnings and Clahsen (2008). L2ers showed effects of the Category Constraint later in reading, and no online evidence of sensitivity to the Structure (i.e., morphological) Constraint. Clahsen et al. (2013) argue that these results are most consistent with the Shallow Structure Hypothesis (SSH; e.g., Clahsen, Felser, Neubauer, & Silva, 2010), which predicts a superficial analysis of morphologically complex forms by nonnative speakers, and in particular for inflections vs. derivations (Neubauer & Clahsen, 2009; Silva & Clahsen, 2008). Clahsen and colleagues argue for this split on the basis of theoretical distinctions, such as whether the output of the process is a word form within the same lexical entry as the base (inflection), or a new lexical entry (derivation; e.g., Anderson,

1992). Thus, since derivation is assumed to draw from both lexical and grammatical mechanisms (Clahsen et al., 2003), nonnative speakers' failure to resort to the latter would explain why regular inflections do not produce priming effects in this population, while derivations only half-prime their stems. Crucially, Clahsen and colleagues only make this claim for productive, semantically transparent derivations. Following the discussion in Clahsen et al. (2003), where productive derivation is said to share features with both regular and irregular inflection, differences should be expected for nonnative speakers between transparent and opaque derivations and/or compounds, since only the former would entail the online recruitment of grammatical processing mechanisms.

While some of the most relevant data on native morphological processing comes from studies on compounding, this word-formation process is underrepresented in the nonnative literature. To our knowledge, only one other published study (De Cat, Klepousniotou, & Baayen, 2015) has focused on compound representation and processing in nonnative speakers. In two (overtly) primed lexical decision experiments testing noun-noun or root compounds in English, De Cat and colleagues analysed the performance of native speakers as compared to that of two different groups of intermediate to advanced L2 learners which differed in their first language: Spanish and German. (A)symmetries in the head directionality patterns of these items—German and English root compounds are head-final, while Spanish presents head-initial compounds—were exploited to evaluate representation vs. processing accounts of L2 morphological performance. The first experiment found that all groups displayed the same accuracy rates for compounds with a licit constituent order (i.e., right-headed), while differences emerged for the reversed order condition: the accuracy of Spanish native speakers dropped as a consequence of their over-acceptance of these items. De Cat and colleagues take these two results as indicative that no representational deficits

seem to be in place for these speakers, but L1 influence in processing of the L2 is apparent on the behaviour of the Spanish learners. ERP recordings in the second experiment, a delayed lexical decision task, further support this explanation: the L1 Spanish group showed comparable compound frequency and crossover constituent frequency effects only for items with illicit order in English, which would nevertheless be canonical in their L1. German native speakers, on the other hand, displayed an extreme sensitivity to violations of word order, which coincides in their native and nonnative languages.

The study presented here differs in scope, aims and methods with De Cat et al. (2015). Their main research question had to do with the influence of the L1 on the processing of a second language, for which they exploited differences in compound formation between the three languages. Their methodology was tailored to address this issue, and is not, in our opinion, optimal to test constituency effects. At a SOA of 100 ms, prime processing has time to activate semantic features that might act as a confound of purely morphological (i.e., structural) effects. Our study seeks to probe for the existence of morphological structure in the nonnative lexicon, employing a masked constituent priming paradigm with native ($n = 71$) and lower-advanced nonnative speakers of English ($n = 78$; L1 Spanish). The experiment was designed to assess whether the lexical entries of synthetic NV-er compounds (e.g., *cheerleader*) are specified for constituent structure. Predictions were as follows: natives were expected to show reliable priming effects for both constituents (e.g., *cheer – CHEERLEADER*; *leader – CHEERLEADER*), replicating a consistent finding in the literature (e.g., Duñabeitia et al., 2009; Libben et al., 2003; Shoolman & Andrews, 2003; Zwitserlood, 1994, among others). Nonnative speakers could go either way: if Diependaele, Duñabeitia and colleagues are right, we would see constituent priming effects

comparable to those of native speakers, albeit with longer response times overall. If Clahsen and colleagues are on the right track, then we would see reduced priming effects for nonnative speakers as compared to native speakers beyond and above their longer response times. Note that, although the SSH does not make explicit predictions about compounding, the argument in Clahsen et. al (2013) and Silva and Clahsen (2008) that derivation (unlike inflection) creates new lexical entries also applies to compounds, which would extend their prediction of a reduction of structurally motivated effects in nonnative speakers (as found in Silva & Clahsen, 2008). Importantly, this should be true especially of compounds that are semantically transparent and/or generated productively (as in the case of NV-er compounds in English), where the SSH predicts grammatical processing to play a larger role.

Method

Participants

A total of 149 undergraduate students (86 females), aged 19-30 (mean = 20.5; median = 20), took part in the experiment. Of these, 71 were native speakers of English (54 females; mean age = 20.6; median = 20; range: 19-30), students of various majors at an American university, who took part in the experiment in exchange for course credit or economic compensation. Due to the controlled presence of a number of Spanish-English cognates in the materials (see the following section), we ensured that none of them had significant knowledge of Spanish by controlling their linguistic background and other potential factors of exposure (e.g., TV) through an interview and a standard questionnaire as described below.

(INSERT TABLE 1 AROUND HERE)

The L2 group was recruited at a Colombian university, and consisted of 78 undergraduates (32 females; mean age = 20.3; median = 20; range = 19-26) then at the 5th to 8th semesters of the degree in English Studies, taught exclusively in English. As part of course requirements, they had taken standardised tests (Oxford Placement Test) around one month prior to the experiment to ensure that they had an upper-intermediate to advanced knowledge of English. They were all native speakers of Spanish, born and raised in Colombia, and had received formal instruction in English since age 10. All participants had normal or corrected-to-normal vision, no history of brain impairment or learning disabilities, and were naïve with regard to the purpose of the experiment.

Materials

Forty (40) English synthetic NV-er compounds (e.g., *cheerleader*) were used as targets. These were matched for total length (mean = 10.19 characters, standard deviation (*SD*) = 0.93, range: 8–12), constituent length (first constituent mean length = 4.15, *SD* = 0.70, range: 3–5; second constituent mean length = 6.03, *SD* = 0.62, range: 5–7), and frequency[2] (mean = 1.99 occurrences per million (o.p.m.); *SD* = 2.26, range: 0.5–10.58). All compounds had their highest number of o.p.m. in solid (i.e., un-split) form. An original list of 43 compounds was used to conduct a norming study among 62 Spanish-speaking participants—a different group of students from the same university as the L2 participants—who were also upper-intermediate to advanced in English. Participants provided a definition for each word and a confidence level in their answer ranging from 1 (minimum) to 5 (maximum). Only compounds which had more than 50% overall score and an average confidence level of 3 points or higher were used in the study. Eleven (11) out of the 40 target compounds had one or two constituents that were cognates with Spanish words. In order to keep this aspect controlled, the five

counterbalancing lists were arranged so that each contained an equal number (7) of cognates across conditions.

For every trial, the target compounds could be preceded by a prime from either one of five conditions:

(1)   1st morphological condition or first constituent; M1 (e.g. *fund – FUNDRAISER*).

(2)   2nd morphological condition or second constituent; M2 (e.g., *raiser – FUNDRAISER*).

(3)   1st orthographic condition; O1 (e.g., *funk – FUNDRAISER*).

(4)   2nd orthographic condition; O2 (e.g., *raisin – FUNDRAISER*).

(5)   Unrelated condition; UN (e.g., *cool – FUNDRAISER*).

The unrelated condition, in which the compound was preceded by a monomorphemic word that had no semantic, morphological or orthographic relationship with it, served as the baseline to which we compared all other prime types. The function of the two orthographic conditions was to control for possible form effects that were orthographic rather than morphological in nature.[3] We used word primes that overlapped orthographically, but not morphologically, with either the first or the second constituent of the compound, matching them in length to the corresponding morphological condition (O1 mean length = 4.45, SD = 0.81; O2 mean length = 6.03, SD = 0.66). Overlap percentages between the morphological and orthographic conditions (M1 − O1: mean = 67.25%, SD = 5.90%; M2 − O2: mean = 54.05%, SD = 7.93%) were calculated according to the SOLAR model of visual word recognition (Davis, 1999, 2006), using *N-Watch* (Davis, 2005).

To make lexical decision possible, a list of 40 pseudowords was generated from the original list of targets by means of letter substitution, using the *Wuggy* software

(Keuleers & Brysbaert, 2010). All items in this category respected the phonotactics of English[4]. The mean overlap percentage was 61.05% (SD = 11.27%). A separate list of 40 monomorphemic words, matched in length with the rest of the primes, were used as primes for the non-word targets (e.g., *brand – FUREPAISER*) to avoid repetition effects. Each participant saw the 40 target compounds only once, preceded by a different set of 8 primes from each of the five conditions. A total of 5 counterbalanced lists were created, which were then administered at random to a roughly equal number of participants—around 16 per list in the non-native group, 13 in the case of native speakers. Each list was filled with an extra 80 targets, made up of N-N compound words (e.g., *snowball*) and derived pseudowords of approximately the same length (mean = 8.45, range: 7–10) and frequency (mean = 13.21 o.p.m., SD = 11.37) as the original compounds. These pseudowords were also preceded by a new list of 40 monomorphemic words, matched in length with the rest of the primes in the set. The relevant NV-er synthetic compound targets thus amounted to 25% of the total consciously processed words.

Procedure

The task was administered on computers running DMDX software (Forster & Forster, 2003), and displayed using CRT monitors with a refresh rate of 80 Hz. Participants sat in a well-lit booth at approximately 50-60 cm from the computer screen. Response buttons were labelled ('YES' and 'NO') on the keyboard, and participants were asked to keep one finger on top of each at all times. The software presented the stimuli of the corresponding list in five blocks of 32 items. Conditions were balanced, but the order of presentation of both the blocks and the items within was randomised for each participant. The task started with a brief set of instructions and a practice session of 10 trials. All visual stimuli were presented in the centre of the screen, in dark red (RGB

#187001001) 40-point size Courier New (a mono-spaced font) over a light grey background (RGB #211211211) to provide a balanced contrast. Each trial began with a 500 ms presentation of a forward mask made up of a row of hash signs (e.g., '#####'), matching the maximum length of the corresponding prime category. Immediately afterwards, the prime—in lowercase letters—was displayed for 62 ms (5 refresh cycles of a 80 Hz monitor), followed by the target –in uppercase. Response latencies were measured from the onset of the target word. Participants were instructed to press the corresponding button (labelled according to handedness) as quickly and accurately as possible, indicating whether the target was an existing word in English or not. Upon response, or after a timeout period of 2500 ms, the target disappeared followed by the forward mask of the upcoming trial. After 32 trials, the software paused and asked the participant to continue on to the next block when ready to do so by pressing the spacebar.

(INSERT FIGURE 1 ABOUT HERE)

Participants were debriefed about the procedure, in order to assess perception of the primes. No participants in the L2 group ($n = 78$) reported having noticed the masked primes. As for the native speaker group, 3 out of 66 participants reported having noticed 'an occasional flash' or 'some letter shuffling', but were unable to tell whether it was a word or a random sequence of letters. Participants were asked to provide some demographics and linguistic background information by completing an English or Spanish version of the *LEAP-Q* (Marian, Blumenfeld, & Kaushanskaya, 2007). In the case of the nonnative speakers, a post-hoc test was administered to further control for knowledge of both primes and targets. Participants were given all primes and targets appearing in their list of the experiment, providing either a translation into Spanish or a

definition in English, along with an indication of their level of confidence in their response (1 to 5). No results from these measures compromise the results of the experiment.

Analyses

The data were prepared by rejecting all potential involuntary inputs and identifying RT outliers within the observations pertaining to correct responses. The criterion to exclude involuntary inputs was set to 200 ms, finding no cases. Following an inverse transformation of the response times to reduce the typical long-tailed positive skewness in the distribution (-1000/RT; Kliegl, Masson, & Richter, 2010; see Diependaele et al., 2011, for a similar approach), outliers were defined as any values lower than $Q_1 - 3 \times IQR$ or higher than $Q_3 + 3 \times IQR$, where Q_1 and Q_3 are the first and third quartiles, respectively, and *IQR* is the inter-quartile range. 11 cases were identified within the full sample of 5346 correct-response observations (which themselves constitute the 89.7% of the total 5960 observations for the word stimuli). These potentially influential data points were monitored, and all statistical analyses were performed with and without the observations. No significant differences emerged; therefore, all results reported belong to the full sample.

All processing, plotting and analysis of the raw data was performed in *R* (R Core Team, 2013). Non-word information was discarded, and both accuracy rates and response times (RTs) to correct responses of word targets were analysed using linear mixed models, with Item and Subject as crossed random factors (e.g., Baayen, Davidson, & Bates, 2008) and Group and Condition as fixed factors, using the *lme4* package in *R* (Bates, Maechler, Bolker, & Walker, 2013). Random slopes were then allowed for select random factors, to obtain the maximal random structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013). Accuracy data were analysed using a

generalised linear mixed effects (GLME) model with a binomial family, whereas a Gaussian family was used for response times. Logarithmic transformations of whole-word, constituent and prime frequency, as well as position-specific morphological family size (e.g., Dijkstra, Moscoso del Prado Martín, Schulpen, Schreuder, & Baayen, 2005; Moscoso del Prado Martín et al., 2005) were included in the process of iterative model building. A separate norming study with a sample of 40 native speakers, different from our experimental population, provided non-discrete transparency ratings for the target compounds ranging from 1 to 4. This information was also incorporated into the models.

Results

Table 2 contains a summary of the accuracy scores and response latencies for each condition and group in the experiment.

(INSERT TABLE 2 ABOUT HERE)

Visual inspection of the models' residuals revealed no obvious signs of heteroscedasticity. Collinearity among predictors was not an issue (highest correlation between fixed effects: whole-word–first constituent frequency = 0.135 [RTs], 0.091 [Accuracy]). The final random structure of the accuracy model presented random intercepts for Subjects and Items, while that of the RT model included, besides these intercepts, random slopes for Subjects and Items within Condition as well as for Items within Group—because the stimuli could have affected each group differently—and Subject within whole-word frequency—because different amount of exposure to a particular word can result in different frequency effects. Table 3 presents the coefficients and associated statistics of the accuracy and RT models.

(INSERT TABLE 3 ABOUT HERE)

In the accuracy analysis, Group, $\chi^2(1) = 7.21$, $p < .001$, displayed a highly significant effect, while Condition did not, $\chi^2(4) = 4.34$, $p = .36$. There was also no interaction between the two, $\chi^2(4) = 4.37$, $p = .36$. The RT analysis found significant effects of both Group, $\chi^2(1) = 97.83$, $p < .0001$, and Condition, $\chi^2(4) = 97.55$, $p < .0001$, with no interaction, $\chi^2(4) = 9.46$, $p = .051$. Post-hoc comparisons of means per Condition using Tukey HSD tests confirmed that significant facilitation was shown only by the morphological, M1 − UN: z = −5.92, $p < .0001$; M2 − UN: z = −6.72, $p < .0001$, but not the orthographic conditions, O1 − UN: z = −1.48, $p = .58$; O2 − UN: z = −0.95, $p = .82$ (cf. p values in Table 3). The first and second constituents, which barely differed in their effects, M1 − M2: z = 1.08, $p = .81$, enhanced responses as compared not only to the baseline unrelated condition, but also to both orthographic prime types, O1 − M1: z = 4.37, $p < .0005$; O2 − M1: z = 4.97, $p < .0001$; O1 − M2: z = 5.66, $p < .0001$; O2 − M2: z = 6.25, $p < .0001$, which themselves had comparable non-significant effects, O1 − O2: z = −0.66, $p = .96$.

Additionally, we found that an increase in semantic transparency interacted positively with the effect of Group, which indicates that the gap between native and nonnative speakers widened as a function of transparency. This factor enhanced responses in the native speaker group while not significantly affecting the nonnative speakers.[5] Crucially, however, this factor did not interact with Condition in accuracy, $\chi^2(5) = 3.34$, $p = .65$, or RTs, $\chi^2(4) = 1.49$, $p = .83$, nor did it enter into a three-way interaction with Condition and Group, $\chi^2(8) = 2.82$, $p = .95$, which has important implications for the role of meaning in the processing of these compounds.

As for frequency measures, whole-word, Accuracy: $\chi^2(1) = 4.75$, $p < .05$; RTs: $\chi^2(1) = 8.61$, $p < .01$, and first constituent frequency, Accuracy: $\chi^2(1) = 16.25$, $p <$

.0001; RTs: $\chi^2(1) = 13.61$, $p < .001$, had an overall facilitating effect on both dependent variables; in general terms, higher frequencies meant faster and more accurate responses (see also Shoolman & Andrews, 2003). In the error data, an interaction with the Group factor showed that, for nonnative speakers, the facilitation produced by higher first constituent frequencies was more important than for native speakers, $\chi^2(1) = 12.59$, $p < .001$. Overall prime frequency was only a significant predictor of Accuracy, $\chi^2(1) = 5.42$, $p < .05$, but not of RTs, $\chi^2(1) = 0.0016$, $p = .97$. This effect on accuracy, which did not differ between the groups, $\chi^2(1) = 0.04$, $p = .85$, or interact with Condition, $\chi^2(4) = 8.99$, $p = .06$, had a negative direction (i.e., accuracy decreased at higher prime frequencies), as did the non-significant numerical trend on the RTs, most likely reflecting some form of prime-target interference. Finally, the accuracy scores of nonnative speakers were affected by the compound's right morphological family size, Group by RMFS interaction: $\chi^2(2) = 7.80$, $p < .05$. This suggests that the existence of a higher number of compounds with the same right constituent improved the L2 speakers' recognition of the targets.

Furthermore, a complex interaction involving whole-word and first constituent frequencies as well as Group and left morphological family size emerged in the response time analysis. However, this kind of interactions entail extremely difficult interpretations that we will not venture for two main reasons. The first is that some of these factors (particularly whole-word frequency but also, to the extent possible, constituent frequency) were controlled in the design of materials, and so are not likely to present a distribution that is as wide as expected in multiple regression studies. The second is that response times are an end-point measure, and as such are difficult to interpret in the same way as it is usually done with methods of high temporal

resolution—where, for example, considerations of left-to-right processing have specific correlates in the data.

Discussion

Constituent priming

The results presented in this study provide strong evidence of constituent priming in English compound processing, both for native and nonnative speakers. Responses to target compounds were enhanced by previous presentation of their constituents in isolation, irrespective of position (first or second). Notably, all significant effects were found in response latencies, whereas accuracy scores seemed not to be affected by experimental manipulations. This is not a surprising finding, nor is it unprecedented in the literature, where responses to unrelated primes have typically been found to be numerically less accurate, but not always significantly so (Frost et al., 2005; Rastle et al., 2004; but see also Diependaele et al., 2011; Forster & Azuma, 2000 for significant differences). The error data of our nonnative speakers is the most prototypical, with 3-4% less accuracy in the unrelated condition. However, unlike some of the previous studies, we do not have a Relatedness variable nested within Condition, but rather an unrelated condition that serves as the baseline for all conditions in every target. This is a design advantage that allows us to control for between-items variability, but makes the (non-central) relatedness effect, broken down into different prime types, less directly observable.

Morphological vs. orthographic effects

The orthographic conditions (e.g., *funk – FUNDRAISER*; *raisin – FUNDRAISER*), in contrast, did not enhance response times, which we attribute to the fact that their recognition does not entail activation of the morphemic structure of the target

compounds. It should be noted, however, that our orthographic control condition is not strictly comparable to the one used in studies of derivational morphology (e.g., *brothel – BROTH*, where decomposition of the prime is non-felicitous, vs. the opaque morphological condition *brother – BROTH*, where the process returns two valid morphemes), since ours did not include a 100% non-morphological form overlap. Nonetheless, it would still be very difficult to provide a purely orthographic account of the effects observed. As reported in the methods subsection above, the mean overlap between the morphological and the orthographic conditions was 67.25% for the first constituent and 54.05% for the second. When considered with respect to the target compounds, these percentages would be 27.34% and 32.07% respectively. This means that the possible orthographic effect would have to plummet from $z = -5.92$, $p < .0001$, to $z = -1.48$, $p = .58$, and from $z = -6.72$, $p < .0001$, to $z = -0.95$, $p = .82$, a rather steep decrease for a difference in overlap of 13 and 27 perceptual points, respectively. Recall that, besides being significantly faster with respect to the unrelated baseline, both morphological conditions were also significantly faster than each of the orthographic conditions. In order to test directly whether differences in degree of orthographic overlap can explain the effects at hand, we conducted new analyses in which the overlap percentage between each prime and its target is included as a factor in the model. Results reveal that this was not a significant factor affecting response times, neither in isolation, $\chi^2(1) = 0.078$, $p = .79$, nor in interaction with Condition, $\chi^2(5) = 6.95$, $p = .22$, or Group, $\chi^2(2) = 1.44$, $p = .49$. These statistical analyses demonstrate that the priming effect is not significantly determined by prime-target orthographic overlap. Given these results, we believe that the weight of the evidence still points to a morphological (as opposed to merely orthographic and/or semantic) locus of effects as the most likely account.

Access and representation

Our results, especially those in the native speaker group, are compatible with some of

the current models of lexical access in visual word processing, as long as these include

morphologically structured representations even for non-transparent and/or highly

lexicalised compounds. The finding that semantic transparency did not interact with the

Condition factor in our response time analyses is particularly illuminating. The

morphological effect obtains irrespective of constituent-to-compound transparency, for

both speaker populations—recall that there was no triple interaction with the Group

factor either. This does not mean that semantic transparency affects both populations

equally (see discussion below), but it speaks to the robustness of the morphological

effect in the face of variable semantic contexts. As for processing mechanisms, our

results favour the view of some stage of pre- or post-lexical decomposition. While this

last aspect is not directly tested by the original design, further analyses revealed

significant effects and interactions of first constituent frequency and family size over

and above those produced by the type of prime-target relationship in each condition.

The interaction of these measures with whole-word frequency may be an indication that

the first constituent acts as a dominant index of access to compound lexical

representations, either early in decomposition or in a later compositional stage (cf.

Inhoff, Starr, Solomon, & Placke, 2008; Juhasz et al., 2003).

In terms of representation, at least some version of most full-parsing (Stockall &

Marantz, 2006; Taft & Nillsen, 2013) and dual/multiple-route models (Kuperman,

Schreuder, Bertram, & Baayen, 2009; MacGregor & Shtyrov, 2013) can accommodate

our results. The first group would expect access to the full form only after the

constituents have been activated through segmentation. Since the left constituent is

recognised first in serial order (cf. Andrews, Miller, & Rayner, 2004; Kuperman et al.,

2009), its properties come into play early—hence the main effect displayed by its frequency. Once the lemma has been activated, a large family size (or a high family frequency; see Taft, 2004) may delay composition, since it will take longer to find the current combination among the many stored or possible competitors. Whole-word frequency becomes particularly relevant, with more frequent compounds having more salient full form lemmas.

Dual-route models predict the (de)compositional route to dominate when processing long or low-frequency complex words—e.g., most of our targets—since the full form would take longer to retrieve. High-frequency left constituents would be recognised and activated faster, further speeding the decompositional route. With a large family size the initial advantage would be counteracted to some extent, slowing down the composition process and shortening the distance with the whole-word route. A simultaneous, multiple-route explanation is also possible: if, as Kuperman et al. (2009) propose, whole-word frequency effects may already obtain before both constituents have been fixated, then a high-frequency first constituent should contribute to speeding up the processing of low-frequency compounds by pre-activating their full forms, any enhancement being attenuated if the number of compounds to activate is large. The absence of observable effects—either in frequency or family size—for the second constituent could then be taken as evidence that this is a matter of activation through partial recognition of whole-word forms, rather than composition per se (but see Fiorentino et al., 2014; Ji et al., 2011, for electrophysiological and behavioural evidence suggesting a vital role of the composition stage). To adjudicate between these theories, future work should gain temporal insight into the whole-word by first constituent frequency interaction. In validation of full-parsing models, first constituent frequency should have an earlier detectable effect on a high temporal resolution measure (e.g.,

ERP, MEG), followed by some sort of signature for the whole-word frequency effect. Under a multiple-route account, alternatively, whole-word and first constituent frequency correlates should be simultaneous and therefore virtually indistinguishable.

One might argue that the absence of a significant interaction effect between Condition and the two morphological family size measures goes against the predictions of theories involving morphological decomposition: if the first or second constituent is activated (as a prime) in the morphological conditions, we would expect to see the observed priming effect vary as a function of that constituent's morphological family size, since a large MFS should either delay—by spreading activation among a larger number of combinatorial alternatives—or enhance—by making the constituent more readily available itself—response times to target compounds. However, we did not compute morphological family sizes independently of position. While the effects of morphological family size in monomorphemes are cumulative with respect to different positions of the morpheme within complex words (e.g., both {fund + ing} and {re + fund} count towards the total morphological family size of *fund*), when processing a compound word serial-order considerations come into play. All morphological neighbours of the compound in which the left constituent appears as the second or third constituent (and vice versa) will most likely not be considered for predictive processing. For that reason, we identified the number of entries corresponding to complex lexical items that contained the first and second constituents of each target compound in initial or final position, respectively, in SUBTLEX$_{US}$, and called these variables Left and Right morphological family size (hence LMFS and RMFS). Therefore, our count can only be considered a fraction (in various proportions) of the total family frequency of the constituent, as it included only those entries in which this word occupied the same position as it did in the compound—e.g., for *babysitter*, it would include *babyface*, but

not *crybaby*. This aspect, which would most strongly affect the count for the first constituent—deverbal nouns like *maker* are rarely if ever in initial position—should be taken into account when considering our data in light of such predictions.

Native vs. nonnative morphological processing

Crucially, native and nonnative speakers barely differed in anything other than total accuracy and average response time, with longer latencies in the nonnative group most likely attributable to their lower proficiency as compared to native speakers. Likewise, a larger reliance on form processing (see, e.g., Diependaele et al., 2011) was not found in nonnative speakers: they were just as insensitive to purely orthographic priming as their native peers. There were, however, some subtle differences when we looked at the item characteristics of our stimuli.

First, semantic transparency seemed to affect exclusively the native group—and this only mildly, $p < .047$. Note that the effect is found across all levels of the Condition factor, with which it did not interact. This finding can thus not be explained by models assuming that the lexical representations of semantically opaque complex words must be accessed through their whole-word forms (e.g., Marslen-Wilson et al., 1994), but fits nicely into compositional accounts (e.g., Ji et al., 2011), wherein the mismatch between the constituent-based and stored idiosyncratic meanings leads to a conflict that slows down the recognition process. The size or intensity of this conflict becomes a direct function of semantic opacity. This relative prominence of the composition stage in the native group as compared to the L2ers seems to indicate a larger role of morphological structure in the first group. While this could be interpreted in line of Clahsen and colleagues' SSH, a direct validation of the theory would involve semantic transparency acting as negative predictor of morphological priming in the L2 group. Our data show a difference between the groups, but this is grounded on a positive effect of semantic

transparency on native speakers rather than a hampering effect on non-natives. Moreover, the non-significant effect of transparency in the L2 group points in the opposite direction: if anything, more transparent compounds received faster responses. Another way to look at this result is to consider that the difference may simply be quantitative, not qualitative: if constituent and whole-word meanings are activated more rapidly by native speakers, a conflict between stored and computed meanings will take place for opaque compounds (see, e.g., Ji et al., 2011). This effect would be attenuated, or even absent, in the case of nonnative speakers, if semantic features take longer to be activated. Nevertheless, note that nonnative speakers gave more accurate responses to compounds with larger RMFSs, an effect that could be equally attributable to composition, and which was not found in the L1 group.[6]

The second difference between the two groups was found in the size, but not the direction, of an interaction effect: the accuracy and speed of responses significantly increased with higher first constituent frequencies in both groups, but in accuracy this effect was significantly larger for the nonnative group. As with the higher accuracy rates with larger RMFSs or the lower accuracy for higher-frequency primes, it is always difficult to offer an absolute explanation on the error data of a lexical decision task, since knowledge of the target word is by far the most stable predictor. It could be the case that, due to time pressure (i.e., the 2500 ms timeout) nonnative speakers are relying on a salient and well known constituent to provide an answer before the compound has been fully processed, which would suggest slower, quantitatively different processing (McDonald, 2006). Similar biases have previously been reported in lexical decision even for native speakers, such that activation of multiple valid words after segmentation might bias participants towards a positive response before assessing the legitimacy of the combination—or, in some cases, its order (Shoolman & Andrews, 2003). While this

is certainly a possibility, it is almost unfalsifiable in our data: our non-word stimuli were not pseudocompounds formed by two existing words, and therefore we do not have insight into the hit/false-alarm ratio of interest here. We cannot compute subject response bias in relation to first constituent frequency, because there were no trials in the experiment in which the correct response to a compound word was a negative one. In any case, it should be noted that some of our non-word stimuli did contain word-like sequences in initial position: ***fear**twheeker* (created from *heartbreaker*), ***plea**rkinner* (from *breadwinner*), ***sham**pstrower* (from *flamethrower*), etc. These sequences do not correspond, in number of characters, to the first constituent of the corresponding compound (e.g., the "first constituent" of *feartwheeker*, if we take *heartbreaker* as reference, would be "feart", not "fear"), but morphological decomposition mechanisms have no a priori way of detecting this. These items are not a majority among our non-word stimuli, but we believe they are numerous enough to prevent the appearance of such strategic effects. If these had taken place, however, their bearing on the overall results would be to inflate accuracy rates and speed up the response times of nonnative speakers, at least for word targets. Yet, Group was a significant factor in both analyses, suggesting that any potential strategic effects have not obscured the differences between native and nonnative speakers. Crucially, such effects would have impacted word targets across the board, with no differential influence on one condition over the other. In this sense, observations in terms of how morphological and orthographic effects may have arisen in the two populations remain unconfounded.

Conclusions

The present study investigated the processing of English compounds by native and nonnative speakers, using a lexical decision task with masked constituent priming. In order to control for a purely orthographic locus of effects, we introduced two conditions

in which prime-target overlap was orthographic but not morphological in nature (e.g., *funk – FUNDRAISER* vs. *fund – FUNDRAISER*). Both experimental groups responded faster to compounds preceded by their first or second constituents as compared to an unrelated baseline (e.g., *coal – FUNDRAISER*), while showing no facilitation from orthographic conditions. These effects were unconditioned by the semantic transparency of the compound, which is problematic for models that conceive of morphological effects as a virtual by-product of the orthography-semantics interface. Main differences in accuracy and response times between the groups do not seem to be attributable to any factors beyond their native language. A further look at the data considering frequency and family size characteristics of the target compounds reveals subtle differences between native and nonnative speakers, compatible with an account in terms of overall slower processing on the part of L2ers.

Our results provide strong evidence of morphologically structured lexical representations in nonnative speakers, and suggest that they use this information in ways virtually identical to native speakers, at least at intermediate to advanced levels of proficiency. These data are difficult to account for by theories postulating inevitable representational differences (see Meisel, 2011, for review), as well as those that propose fundamental differences in processing, whether or not these are linked to representation (Clahsen et al., 2010). Further research should be conducted to determine whether the level of specification of these representations extends beyond morphemic structure. Future cross-sectional studies, in taking up the (methodological) challenge of evaluating the role of proficiency and AoA of the L2 in the development of morphological structure within the lexicon, might help integrate seemingly disparate results and enhance our understanding of the dynamic patterns of morpho-lexical acquisition and processing. Further research should also be conducted to assess the strength and level of

specification of these morphological representations, and more specifically to test the hypothesis that the purported slower processing capacity of nonnative users of a language may itself be sufficient to account for most of the differences observed in morphological processing with respect to native speakers. More widespread use of experimental methodologies with high timescale resolution, such as eye-tracking, ERP and MEG measures (cf. Clahsen et al., 2013; Kuperman, 2013; Fiorentino et al., 2014; De Cat et al., 2015), may provide invaluable insight into the time course of access to morphological information in nonnative language processing, and allow us to further inspect whether any qualitative differences exist between native and nonnative speakers in their use and representation of morphology. We believe that most currently available evidence, including the data from the present experiment, points to the contrary.

Acknowledgements

References

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, *16*(1-2), 285–311.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. Vienna, Austria: R Foundation for Statistical Computing.

Bijeljac-Babic, R., Biardeau, a, & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Mmeory and Cognition*, *25*(4), 447–457.

Bozic, M., Marslen-Wilson, W. D., Stamatakis, E. a, Davis, M. H., & Tyler, L. K. (2007). Differentiating morphology, form, and meaning: neural correlates of morphological complexity. *Journal of Cognitive Neuroscience*, *19*(9), 1464–75.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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), 977–90.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what , when , where , and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98.

Chateau, D., Knudsen, E. V., & Jared, D. (2002). Masked Priming of Prefixes and the Influence of Spelling–Meaning Consistency. *Brain and Language*, *81*(1-3), 587–600.

Clahsen, H., Balkhair, L., Schutter, J.-S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research*, *29*(1), 7–31.

Clahsen, H., Felser, C., Neubauer, K., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, *60*(1), 21–43.

Clahsen, H., Sonnenstuhl, I., & Blevins, J. P. (2003). Derivational morphology in the German mental lexicon : A Dual Mechanism account. In *Morphological structure in language processing* (pp. 125–155).

Cunnings, I., & Clahsen, H. (2008). The time-course of morphological constraints: A study of plurals inside derived words. *The Mental Lexicon*, *3*, 149–175.

Davis, C. J. (1999). *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition*. University of New South Wales.

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*, 65–70.

Davis, C. J. (2006). Orthographic input coding: A review of behavioural data and current models. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 180–206). Psychology Press.

De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology*, *6*(77), 1–17.

Devlin, J. T., Jamison, H. L., Matthews, P. M., & Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proceedings of the National Academy of Sciences*, *101*(41), 14984–14988.

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., & Baayen, R. H. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, *20*(1-2), 7–41.

Duñabeitia, J. A., Dimitropoulou, M., Morris, J., & Diependaele, K. (2013). The role of form in morphological priming: Evidence from bilinguals. *Language and Cognitive Processes*, *28*(7), 967–987.

Duñabeitia, J. A., Laka, I., Perea, M., & Carreiras, M. (2009). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *European Journal of Cognitive Psychology*, *21*(4), 615–640.

Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Durđević, D. F., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, *13*(2), 119–135.

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-

orthographic: a violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *16*(4), 684–91.

Fiorentino, R., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon*, *4*(2), 159–193.

Fiorentino, R., Naito-billen, Y., Bost, J., & Fund-Reznicek, E. (2014). Electrophysiological evidence for the morpheme-based combinatoric processing of English compounds. *Cognitive Neuropsychology*, *31*(1-2), 123–146.

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, *22*(7), 953–1000.

Forster, K. I., & Azuma, T. (2000). Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, *15*(4-5), 539–561.

Forster, K. I., & Davis, C. J. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 680–698.

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–24.

Frisson, S., Bélanger, N. N., & Rayner, K. (2014). Phonological and Orthographic Overlap Effects in Fast and Masked Priming. *Quarterly Journal of Experimental Psychology*, *18*(9), 1199–1216.

Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: principles of lexical organization in a given language. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1293–1326.

Grainger, J., Granier, J.-P., Farioli, F., Van Assche, E., & van Heuven, W. J. B. (2006). Letter position information and printed word perception: the relative-position priming constraint. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(4), 865–84.

Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The Time Course of and Phonological Orthographic Code Activation. *Psychological Science*, *17*(12), 1021–1026.

Hyönä, J. (2012). The role of visual acuity and segmentation cues in compound word identification. *Frontiers in Psychology*, *3*(1), 1–8.

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory & Cognition*, *36*(3), 675–687.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406–430.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology (London, England : 1953)*, *94*(Pt 2), 223–44.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–33.

Ktori, M., Midgley, K., Holcomb, P. J., & Grainger, J. (2015). An ERP investigation of orthographic priming with superset primes. *Brain Research*, *1594*, 233–244.

Kuperman, V. (2013). Accentuate the positive: semantic access in English compounds. *Frontiers in Psychology*, *4*(203), 1–10.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895.

Lee, H. W., Rayner, K., & Pollatsek, a. (1999). The time course of phonological, semantic, and orthographic coding in reading: evidence from the fast-priming technique. *Psychonomic Bulletin & Review*, *6*(4), 624–634.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: the role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50–64.

MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, *126*(2), 217–229.

Marelli, M., Zonca, G., Contardi, A., & Luzzatti, C. (2014). The representation of compound headedness in the mental lexicon: A picture naming study in aphasia. *Cognitive Neuropsychology*, *31*(1-2), 26–39.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience And Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940–967.

Marslen-Wilson, W. D. (2007). Morphological processes in language comprehension. In G. Gaskel (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 175–193). Oxford: Oxford University Press.

Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B*, *362*, 823–836.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401.

Meisel, J. (2011). *First and Second Language Acquisition: Parallels and Differences*. Cambridge: Cambridge University Press.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language*, *53*(4), 496–512.

Neubauer, K., & Clahsen, H. (2009). Decomposition of Inflected Words in a Second Language. *Studies in Second Language Acquisition*, *31*(3), 403–435.

R Core Team, T. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7-8), 942–971.

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*(6), 1090–1098.

Robert, C., & Mathey, S. (2012). The Effect of Prime Duration in Masked Orthographic Priming Depends on Neighborhood Distribution. *Language and Speech*, *55*, 249–262.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology Section A*, *42*(3), 529–567.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *TRENDS in Cognitive Sciences*, *4*, 353–361.

Semenza, C., & Luzzatti, C. (2014). Combining words in the brain: The processing of compound words. Introduction to the special issue. *Cognitive Neuropsychology*, *31*(1-2), 1–7.

Shoolman, N., & Andrews, S. (2003). Racehorses, Reindeer and Sparrows: Using masked priming to investigate morphological influences on compound word identification. In S. Kinoshita & Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 241–278). Taylor and Francis.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, *1*, 85–123.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647.

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.

Taft, M., & Nillsen, C. (2013). Morphological decomposition and the transposed-letter (TL) position effect. *Language and Cognitive Processes*, *28*(7), 917–938.

Van Heuven, W. J., Dijkstra, T., Grainger, J., & Schriefers, H. (2001). Shared neighborhood effects in masked orthographic priming. *Psychonomic Bulletin & Review*, *8*(1), 96–101.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compound words. *Language and Cognitive Processes*, *9*, 341–368.

Footnotes

[1] Although this last aspect was not directly controlled by the authors, an inspection of the materials in the appendix reveals that their experiment contained transparent (e.g. *lanpostu*, lit. 'job post'), partially opaque (e.g. *hitzaldi*, 'conference talk', lit. 'word time') and fully opaque (e.g. *eskubide*, 'right' –as in 'human rights'–, lit. 'hand way') compounds alike, warranting a certain degree of speculation. However, studies addressing the transparency factor more specifically (e.g. Fiorentino & Fund-Reznicek, 2009) are perhaps better suited to make claims about the morpho-lexical (as opposed to lexico-semantic) nature of constituent priming in compounds.

[2] Lemma frequencies taken from the SUBTLEX$_{US}$ corpus (Brysbaert & New, 2009), containing a total of 51 million words. Only solid (i.e., non-hyphenated, non-fractured) occurrences of the compounds were included in the count.

[3] A *direct* assessment of the claim by proponents of distributed-connectionist models that what seem like morphological effects are, in fact, combined effects of form and meaning overlap would have required a semantic control condition as well, outside our primary concern (morpho-lexical effects).

[4] Even though they were constructed from synthetic NV-er compounds and even retained the (pseudo-) morphological ending *–er*, the non-words in our experiment were not pseudo-compounds in the traditional sense (a nonsensical combination of two otherwise existing words, e.g., *climbsmoke*). They were, however, pseudowords being phonotactically plausible, non-existing letter strings. Please refer to the discussion section for some speculation on how this could have affected nonnative performance. We thank two anonymous reviewers for their comments on this issue.

[5] The interpretation of main effects in the presence of interactions in which they are involved is not always straightforward. In order to assess the independent influence of Group on the dependent variable, we fitted an unconditional model without the interaction terms in which Group appeared. The estimates of that model confirmed the direction and significance of the effect. This is equally applicable to the main effects of whole-word and first constituent frequency.

[6] We thank an anonymous reviewer for suggesting the differential speed of access account. Note that nonnative speakers' higher accuracy on compounds with a large RMFS can be explained in ways not necessarily involving a compositional stage (although still assuming segmentation). Perhaps nonnative speakers take MFS as a measure of lexicality for the whole compound: the more compounds exist with the same second constituent, the more likely it is a valid word. The

fact that we did not find this effect for LMFS may suggest that nonnative speakers are aware of the relative importance of the head constituent, but no claims of a directionality effect can be made without explicitly showing significant differences between both lateral MFSs (instead of between one of them and a baseline or reference level).