Compound words and structure in the lexicon

Robert Fiorentino

Department of Linguistics, University of Kansas, Lawrence, KS, USA

David Poeppel

*Cognitive Neuroscience of Language Laboratory, Department of Linguistics
and Department of Biology, Neuroscience and Cognitive Science Program,
University of Maryland, College Park, MD, USA*

The structure of lexical entries and the status of lexical decomposition remain controversial. In the psycholinguistic literature, one aspect of this debate concerns the psychological reality of the morphological complexity difference between compound words (*teacup*) and single words (*crescent*). The present study investigates morphological decomposition in compound words using visual lexical decision with simultaneous magnetoencephalography (MEG), comparing compounds, single words, and pseudomorphemic foils. The results support an account of lexical processing which includes early decomposition of morphologically complex words into constituents. The behavioural differences suggest internally structured representations for compound words, and the early effects of constituents in the electrophysiological signal support the hypothesis of early morphological parsing. These findings add to a growing literature suggesting that the lexicon includes structured representations, consistent with previous findings supporting early morphological parsing using other tasks. The results do not favour two putative constraints, word length and lexicalisation, on early morphological-structure based computation.

INTRODUCTION

The role of morphological complexity in the representation and processing of compound words and inflectionally or derivationally affixed words is hotly

contested (see Forster, 1988; McQueen & Cutler, 1998; Seidenberg & Gonnerman, 2000; Domínguez, Cuetos, & Segui, 2000; Taft, 1991). The experimental literature on this topic over the last 30 years includes research on inflections, on derivationally complex words, and, to some extent, on compound words, which are widely attested in many languages,[1] are structurally built from two stems, and which may show either transparent (e.g., *teacup*) or opaque (e.g., *bellhop*) semantic relations from parts to whole (analyses in Bauer, 1983; Downing, 1977; Levi, 1978; Spencer, 1991, among others). The research on morphological complexity in the psycholinguistic literature has variously supported both decompositional and non-decompositional accounts, and further, putative effects of decomposition that have been identified are not yet well constrained, for example when they occur in the time course of lexical processing. Ultimately, however, these issues become crucial both from the psycholinguistic and the broader cognitive science perspectives, as the differing viewpoints on compound representation and processing make very different claims on the nature of the representation of linguistic material in the cognitive architecture of language. The aim of the current study is to present a new cognitive neuroscience experimental approach for testing the non-decompositional hypothesis of compounds against the class of decomposition models, adding a neural index of access to constituents to the behavioural measure. We present behavioural and neural evidence for structured representation in the lexicon, which is reflected in the early decompositional processing profile of known compound words. We discuss these findings in the context of the emerging literature on early morphological parsing and other results suggesting abstract structured representations in the lexicon.

Experiments by Taft and Forster (1975) were among the first to use the lexical decision task to investigate the processing of affixed words. Taft and Forster (1976) extended this research to compounding, showing effects of morphological constituency in compounds, which were taken to suggest the online decomposition of complex forms. However, this conception of lexical processing did not go unchallenged. Butterworth (1983) offered a competing analysis of the role of morphological structure in processing, positing a non-decompositional account; this followed from the intuition that full parsing could not work since the idiosyncrasies observed in complex words (such as lack of full productivity for morphological rules) suggested that morphological rules could not drive lexical processing online. In this type of account, words that seem to be morphologically complex are not treated as such;

---

[1] See Bertram & Hyönä (2003) and Janssen, Bi, & Caramazza (2006) for some estimates of the number of compounds in Finnish and English respectively, and Bauer (2001), Plag (1999), Hay & Baayen (2002), among others, for various views on how 'productivity' might be defined and measured quantitatively.

instead, they are stored and processed as whole words. Non-decompositional processing has been claimed for many types of complex word, including words which seem to have been formed by morphological processes such as regular past-tense formation (for some experimental studies supporting this view, at least in part, see Baayen, Dijkstra, & Schreuder, 1997; Manelis & Tharp, 1977; Stemberger & MacWhinney, 1986; Sereno & Jongman, 1997, among others).

Subsequent lexical decision research continued to utilise the basic paradigm of the early experiments such as Taft and Forster (1976), manipulating frequency of compound constituents and looking for differential frequency effects. Two studies on compounds subsequent to Taft and Forster (1975, 1976) which examined the effects of manipulating constituent frequency within compound words are Andrews (1986, Experiments 2 and 3) and Juhasz, Starr, Inhoff, & Placke (2003, Experiment 1). In each case, internal constituent frequency was manipulated and constituent frequency affected reaction time, with higher first or second constituent frequency correlating with response time. Andrews (1986) found consistent effects of constituent frequency. Juhasz et al. (2003) also report constituency effects, noting that first constituent effects were more clear when second constituents were low frequency, suggesting that access to the constituents depended centrally on the properties of the second (head) constituent. Further, although Andrews (1986) found the predicted constituency effects for compounds, the effects for derivationally complex words depended on the stimulus set: the effects were significant only when compound words were part of the stimulus set, leading to the conclusion that decompositional effects, including compound constituent effects, were not prelexical, and were probably controlled rather than automatic. While both results suggest some role for morphological constituency, the computation of constituency and its locus in the time course of lexical processing remain unclear (for additional examples of base/surface differential frequency effects in other types of morphologically complex words, see Colé, Beauvillain, and Segui, 1989; New, Brysbaert, Segui, Ferrand, and Rastle, 2004, among many others).

Eye-tracking

Research on constituent effects has extended beyond lexical decision, for example to studies on eye-tracking.[2] The eye-tracking method has the

---

[2] See Rayner (1998) for a comprehensive review of eye-tracking methodological issues and results. Some challenges for the linking of linguistic computations to component measures in the eye-movement record include accounting for recognition effects that are often distributed over several fixations (e.g., Inhoff et al., 1996) and may be subject to parafoveal preview effects (e.g., Rayner & Pollatsek, 1989; but see Deutsch, Frost, Pollatsek, and Rayner, 2000 for an interesting use of the parafoveal preview benefit to detect morphological decomposition in Hebrew).

advantages of high temporal sensitivity, and thus eye-tracking, like electro-physiology, can potentially play a large role in cross-method research aimed at understanding the role of morphological structure in the time course of lexical processing. Further, unlike lexical decision, eye-tracking has the ability to make measurements during natural reading.[3] Given the potential for mapping computations onto separate time-sensitive components in the eye movement record, eye-tracking is especially relevant for morphological complexity research.

Eye movements have been used to study effects of decomposition of complex words such as compounds (Andrews, Miller, & Rayner, 2004; Bertram & Hyönä, 2003; Inhoff, Briihl, & Schwartz, 1996; Juhasz et al., 2003; Pollatsek & Hyönä, 2005; Pollatsek, Hyönä, & Bertram, 2000, among others). Andrews et al. (2004), for example, recorded eye movements during the reading of English compounds in sentence context, along the lines of work done in Finnish by Pollatsek et al. (2000) and Hyönä & Pollatsek (1998) which showed frequency effects in the reading of Finnish compounds. The results of Andrews et al. (2004) from English suggest some influence of first-constituent frequency on first fixation, and effects of both first and second constituent frequency on gaze duration. Like the earlier studies, whole-word frequency showed an effect on gaze duration and total looking time (in regression analyses on whole-word frequency). In Andrews et al. (2004), these data are taken to reflect a process of *segmentation-through-recognition*, where access to compound words involves processing of both constituent and whole-word representations.[4] Together, these studies point to a role for constituents early in time course, suggesting access to compounds as internally structured representations, inconsistent with a whole-word only approach.

---

[3] A related issue is whether effects found in isolated word tasks like lexical decision would extend to studies of reading. For example, Bertram et al. (2000) found that sentence context seems to play a role in whether decompositional processing is evident for inflected Finnish words carrying the ambiguous suffix *–jA*. Fixation and gaze duration measures, as well as reading time measures in self-paced reading, show base frequency effects (although base frequency effects were typically weak on first fixation durations and reading times on the inflected word, and more robust in gaze duration and in measures on the following word in both methodologies) while in a simple lexical decision task, base frequency effects were not evident. Hyönä, Vainio, and Laine (2002) have found a distinction in the opposite direction – complexity effects in lexical decision but not reading – using Finnish case-marked complex words vs. single words. As regards compounds, constituency effects seem to hold not only in lexical decision tasks but also in eye-tracking studies of compound processing in sentence context (see Pollatsek & Hyönä, 2005 [Finnish], Andrews et al., 2004 [English], among others).

[4] Juhasz et al. (2003) found similar constituent effects for English compounds in naming, lexical decision, and eye movements, finding some effect of first constituent frequency in naming, lexical decision, and first fixations, but more robust second constituent effects in the two behavioural measures and in gaze durations.

Priming

Priming studies have also been used to assess the morphological representation of complex words. These experiments have generally been focused on dissociating the contributions of formal overlap, morphological overlap, and semantic relatedness in the priming of morphologically structured (or pseudo-morphemically structured) complex words and their constituents. These experiments have often relied on delayed repetition priming tasks and cross-modal priming (e.g., Marslen-Wilson, Tyler, Waksler, & Older, 1994; Monsell, 1985). Marslen-Wilson et al. (1994) showed, using a cross-modal repetition priming task, that semantically transparent derived forms showed priming effects, regardless of phonological transparency, although semantically opaque forms did not show priming effects, behaving instead like monomorphemic words (see also Longtin, Segui, & Hallé, 2003, Experiment II, among others).

However, cross-modal priming may be sensitive to semantic factors that come into play *subsequent* to morphological decomposition. For example, the results of Marslen-Wilson et al. (1994), suggested that opaque derived words are monomorphemic in lexical entry since they do not prime in the cross-modal paradigm.[5] Whether a conclusion such as the latter is true of the morphological level is a question that would be better addressed taking the results in context of other tasks which may help to further specify at which level transparent and opaque words differ. One way that researchers have tried, within the priming tradition, is to look at overt immediate repetition priming and masked priming. Among these studies, there are some that have focused specifically on compound words.

---

[5] As one reviewer notes, priming for opaque forms remains evident in cross-modal tasks in Semitic languages. Indeed, Plaut and Gonnerman (2000) explored an account of this phenomenon within a distributed-connectionist model without abstract morphological structure, attributing the difference across languages to the overall level of complex word processing in the language. Plaut and Gonnerman (2000) conducted simulations among artificial languages including a distinction in terms of a 'morphologically rich' language (modelled on Semitic languages), and a 'morphologically poor' language (modelled on English). The intuition behind these simulations is that the morphological properties of a language, such as having a high number of morphologically complex forms, leads to parsing of opaque forms in the morphologically rich, but not the morphologically poor language. However, at least two points are relevant here: (1) the 'morphologically poor' languages like English show robust morphological priming regardless of semantic transparency in masked priming tasks, contrary to the predictions of the typological explanation (see Feldman, 2000; Fiorentino, 2006; Rastle et al., 2000, 2004) and (2) it has been proposed that relatively more persistent priming in Semitic languages may be the result of the relatively lower level of formal overlap compared with English, thus reducing form-level inhibition which might otherwise counter morphological-level facilitation (see Forster, 1998).

Zwitserlood (1994), for example, used the immediate constituent-repetition priming and semantic priming paradigms to explore the processing of semantically transparent, partially transparent, and opaque compounds in Dutch. The results of the two experiments reported there show constituent priming by compound words regardless of semantic transparency. Significant priming was found both for transparent prime-target pairs, such as *kerkorgel – orgel* (gloss: *church organ – organ*) and for opaque pairs, such as *klokhuis – huis* (lit. gloss of prime: *clockhouse*, meaning: *apple – house*). On the other hand, there was no priming for targets with only orthographic overlap, but not morphological constituency, such as *kerstfeest – kers* (gloss: *Christmas – cherry*). When testing the priming of semantic relatives of the target constituents, only the totally and partially transparent items showed significant priming. The results on the partially transparent items contrast with Sandra (1990) who did not find semantic priming from the opaque constituent of the partially transparent compounds.[6] Nevertheless, the results are suggestive of morphological-level complexity for both transparent and opaque compounds at some level, and suggest a difference among morphological and semantic relatedness.

Recently, masked priming (see Forster, 1999 for a recent discussion) has yielded interesting results regarding the processing of morphologically complex words, mainly focusing on derivational morphology (Frost, Forster, & Deutsch, 1997; Longtin, Segui, & Hallé, 2003; Rastle, Davis, & New, 2004, among others). These studies show that masked prime words with apparent morphological complexity significantly facilitate responses to the apparent constituent targets (e.g., priming of 'apart' by 'apartment'), whereas words with orthographic overlap without apparent morphological constituency do not prime the overlapped word part (e.g., no priming for 'elect' by 'electrode'). Such findings suggest that apparently complex words may be parsed rapidly and automatically into morphological-level constituents; we examine these studies in more detail in the Discussion section below. Shoolman and Andrews (2003) used masked priming to test the effect of constituent priming on compound recognition. This study focused on masked priming of compounds (*bookshelf*), pseudo-structured words (*hammock*) and various types of nonword. The results showed both first and second constituent priming of compounds by their constituents regardless of semantic relatedness. Thus, the results from Zwitserlood (1994) seem to hold

---

[6] Sandra (1990) used what amounts to a longer SOA, as Zwitserlood (1994) notes.

even when the prime-target pairs are not consciously compared, and again, favour a morphological-level explanation.[7]

The patterns emerging from the priming literature suggest a role for morphological constituency which is (a) separable from formal overlap, as the former tends to be facilitative and the latter inhibitory in masked priming tasks, and (b) modulated by semantic relatedness but maybe only at some delay, as constituent priming holds for all kinds of constituent structures in masked priming, while constraints such as semantic transparency are detectable in overt, longer lag (e.g., cross-modal) priming tasks, if at all. These findings suggest a broad decompositional conception of the lexicon.

Morphological processing: Direct comparison method

What is virtually absent in the literature is a method for the direct comparison of words varying in internal structure in lexical decision. One previous dataset in English that did allow for such a comparison was Andrews (1986). While Andrews (1986) reported significant constituency effects in the first constituent position, the potentially interesting direct comparisons with monomorphemic controls available in that study were not significant; high frequency first constituent compounds were numerically, but not significantly faster than monomorphemic controls in both compound experiments (Experiments 2 and 3). Although Andrews (1986) controlled for length and number of syllables, and for frequency as well as possible given the sampling error of corpora at very low frequencies (using the Kučera & Francis, 1967 counts) the mean whole-word frequencies reported are higher for the monomorphemic words (2.8) than for the high- (1.8) or low-frequency first-constituent compound stimuli.

We recalculated the frequencies of these items using a newer, but also larger corpus (Collins Cobuild, 320 million words; for Cobuild resources, see http://www.cobuild.collins.co.uk), and tested the differences statistically (note that all monomorphemic words, but not all compounds, were in this corpus; the raw frequency values of the four missing compounds were replaced with the mean raw frequency for that condition). Log frequencies were also higher in this corpus, as in Kučera and Francis (1967), for the

---

[7] These priming results are not without controversy. As regards methodological concerns, the view that masked priming is relatively insensitive to overt strategic effects compared with overt priming has been challenged, for example by Masson and Bodner (2003) and Masson and Isaak (1999). The latter presents the argument that nonword priming effects suggest a pre-lexical locus of masked priming effects. However, Forster (1998, 1999) and others argue that the findings on nonword priming can be explained in context of masked priming operating at the lexical level. The claim that results from the masked priming of complex words implicates the existence of morphological-level constituency has been challenged from the distributed-connectionist viewpoint (see the Models section below for more discussion).

Perspective on Internal Structure	Whole-word representation		Morphologically Structured Entry		
Example	flagship	crescent	flag	ship	crescent
Log Frequency	.68	.69	1.49	1.95	.69
Length	8	8	4	4	8
Syllabicity	2	2	1	1	2

Figure 1. Stimuli were pairwise matched on overall properties, such that overall frequency, syllabicity, and length were matched; the constituent properties of the compounds were significantly different from those of the whole words.

monomorphemic words than the compound words, $F(3, 56) = 5.223$, $MSE = .214$, $p < .004$).[8] This suggests that it might be more difficult to directly compare the compounds and control monomorphemic words directly in that case, although it says nothing about the constituency effects found among the high- versus low-frequency compound words.

In the current study, we take advantage of the direct comparison of compounds and single words, using carefully matched compounds and single words, differing only in *constituent* properties (see Figure 1). This allows us to test directly for differences among words with hypothesised internal morphological structure and those which are monomorphemic in structure, under controlled conditions. The two types of word (compound and single word) make very different predictions under decompositional vs. non-decompositional viewpoints, and under different articulations of decompositional/dual-route models (see Figure 2).

Models

Non-decompositional models.   The non-decompositional model (Figure 2, model III) predicts no online role for morphological-level constituents. Fundamental aspects of the non-decompositional account persist today (i) in those accounts that propose that full storage of complex structures is pervasive (e.g., Bybee, 1995), and (ii) in accounts which claim no abstract

---

[8] Planned contrasts of compound word versus control single word averaged log frequency were significant ($p < .001$) as were comparisons of the compounds with high frequency first constituents versus control single words, and compounds with low frequency first constituents versus controls ($p < .02$ and $p < .009$, respectively). ANOVA on letter length was also significant, [$F(3, 56) = 5.157$, $p < .004$]. A contrast of the compound versus control average length was significant ($p < .004$) as was a contrast of high first constituent compound versus control length ($p < .002$). Letter length also significantly differed among the control sets [paired t-test, $t(29) = 2.8$, $p < .009$, two-tailed]. Letter length among low first constituent frequency compounds versus controls did not differ significantly ($p < .293$).

representation of words at all. For example, Bybee's (1995) model handles 'constituency' as associative relations among related, separate words, where lexical relatedness is defined as strength of connections in phonological and semantic features. Under this model, so-called constituency effects emerge by the associative activation of related forms, mediated by frequency: so-called complex words with high token frequency have weaker lexical relations, and thus are predicted to show reduced 'constituency' effects (Bybee, 1995).

Distributed-connectionist models (e.g., Seidenberg & Gonnerman, 2000) seek to capture relations among whole-words and constituents without recourse to abstract morphemes or a morphological level of analysis. These approaches instead attribute effects among complex words and their constituents to direct form-meaning overlaps (i.e. overlap in phonology, orthography or semantics), which can be modelled using weighted connections in a connectionist network. Data from tasks such as masked priming has been put forth as a challenge to the distributed-connectionist account, since morphological priming has been shown to persist regardless of formal or semantic overlap (i.e. priming holds for semantically transparent and opaque complex words, although words with only orthographic overlap do not show significant priming; for challenges to the conclusion that these results implicate morphological-level processing, see for example Seidenberg and Gonnerman, 2000; Gonnerman, Seidenberg, and Andersen, 2006; for discussion of another network model without abstract lexical representations, see Elman, 2004, and for a recent description of an approach to morphological complexity which challenges the necessity of morphemes, see Hay and Baayen, 2005).

Whole-word processing survives also in how *known* words are treated in many models, including the *supralexical model* (in which initial access is always via whole-word representation, e.g., Giraudo & Grainger, 2000), and *parallel dual route models* (parallel access to whole-word representations, e.g., Schreuder & Baayen, 1995), both models also incorporating some form of morpheme-level processing. We consider these models in turn below.

Late decomposition models.   The late decomposition model (Figure 2, model II) predicts that constituents are activated subsequent to whole-word access. One example is the *supralexical model*, which claims that initial processing proceeds via whole-word representations, with access to morphological constituents following afterward, and only under some circumstances, such as when the relation among whole word and constituents is semantically transparent (e.g., Giraudo & Grainger, 2000; see Diependaele, Sandra, & Grainger, 2005 for more discussion).

Early decomposition models.   Early decomposition models posit that constituent morphemes are activated early and automatically during

processing of a complex word. *Dual-route* models suggest that both decompositional and whole-word processing routes are available, although various factors influence which route will be successful for a given word form. A variety of dual-route models have been proposed, which posit differences in when whole-word and decompositional pathways will be taken due to factors such as transparency, lexicality, productivity, and frequency.

For example, the Morphological Race Model (MRM) (Schreuder & Baayen, 1995; Baayen et al., 1997; Baayen, 1992, among others) is a parallel dual-route model in which decompositional and whole-word access are deployed in parallel and race, allowing for facilitation from parts to whole and fully decompositional parsing in the case of novel or non-listed words (see Figure 2, Model I). While both the MRM and the Augmented Addressed Morphology model (AAM) (e.g., Caramazza, Laudanna, & Romani, 1988) posit that whole-word and morphological-level representations can be accessed, both tend to assume that whole-word processing will typically be more rapid for known compound words (decomposition requires extra steps in processing under these approaches, which can account for effects of whole-word access prior to access to morphological parts). As for the MRM, Schreuder and Baayen (1995) conclude that lexicalised
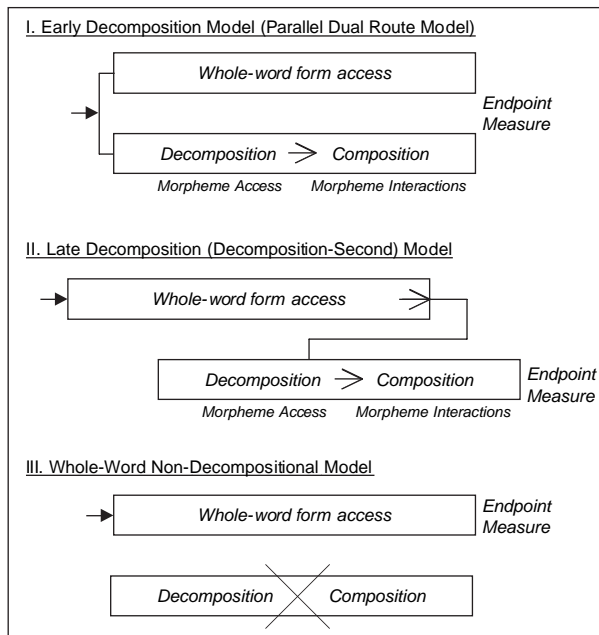
Figure 2. These are three basic models capturing views of the processing of morphologically complex items such as compounds.

compounds are accessed as full-forms, based on a reanalysis of Taft and Forster's high vs. low first-constituent frequency manipulation (Taft & Forster, 1979, Experiment 5).

However, under some variants of a dual-route model, morphemic constituent effects for lexicalised compounds would be predicted. For example, in a parallel dual-route model allowing for activation of constituents to add activation to whole-word representations and vice-versa, and assuming that short, high-frequency constituents are activated early in parsing the compound, this may facilitate access to the whole compound's entry (see for example the Andrews et al., 2004 *segmentation-through-recognition* model, and the parallel dual-route of Pollatsek, Hyönä, and Bertram, 2000). Further, both the MRM and AAM may also be able to capture morpheme-level processing in cases such as high base/low surface frequency mismatches in which morphological stems are of high frequency relative to the whole-word form (for this claim regarding the AAM, see for example Laudanna, Cermele, and Caramazza, 1997).

In contrast, the *full-parsing* approach is explicit in predicting that words are automatically decomposed into constituents, and that all processing is done by the decompositional route (see for example Taft, 2004; Stockall & Marantz, 2006). Access to constituents at an initial stage is only one part of the computation of morphologically complex words along a decompositional route. Some subsequent stage of morpheme combination, which we may term composition, should be a part of the set of computations (see also Baayen & Schreuder, 1995, and Taft, 2004, among others). There are several possibilities regarding what the composition stage might include, under full-parsing, which would affect predictions on the speed of judgements to complex words. Composition may comprise simply 'gluing together' the parts (what this might entail is not clear, nor is whether it is affected by properties like frequency of combination, as speculated by Taft, 2004); alternatively, composition might always involve more costly interpretive combinatorial processes (as is claimed under some versions of the MRM dual-route model, for example).[9] Under a full-parsing model in which combination is costly, we might expect early facilitation of constituent morphemes, but a contrasting delay in response time for both known and novel compounds; however, under a full-parsing model in which morpheme-combination can (sometimes) be rapid, such a model would be able to account for early activation of morphemes and facilitation in response time. Neither variety requires any full-storage of lexicalised compounds.

---

[9] One speculation is that how the effects are manifested may be affected by task, item-set, and other factors thought to influence post-lexical processing.

Electrophysiological component measurement in word recognition

The primary method of research on lexical access has been the lexical decision task. As discussed above, it is clear that lexical decision results are able to speak to important questions regarding the nature of lexical representations, and that the task can also generate results relevant to specifying the locus in time course of various processing stages. Further, lexical decision offers the methodological benefits of a strictly time-locked measurement from word-onset, without look-ahead or other sentential context effects. However, criticism citing the drawbacks of the lexical decision paradigm is not new. A central claim is that lexical decision is not only sensitive to lexical processes but also post-lexical processing (e.g., Balota & Chumbley, 1984; Monsell, Doyle, & Haggard, 1989; Seidenberg, Waters, Sanders, & Langer, 1984). Since effects of constituency are likely to be highly circumscribed in time and to represent a complex set of computations, and since lexical decision is a single response measure taken at the very end stage of processing, it makes sense to add additional dependent variables along the way to decision, to test for the presence of decomposition and the nature of its computation in time course. In this vein, the electrophysiological measure becomes crucial; the interaction of the brain-level effects and behavioural measures can help us narrow in on specific subcomponents of the processing of complex words.

Electrophysiological brain recording measures such as EEG and MEG provide direct measures of neural activity during tasks such as language processing with millisecond temporal resolution. A body of recent MEG research (Cornelissen, Tarkiainen, Helenius, & Salmelin, 2003; Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001; Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998, 1999; Helenius, Salmelin, Service, Connolly, Leinonen, & Lyytinen, 2002; Koyama, Kakigi, Hoshiyama, & Kitamura, 1998; Koyama, Naka, & Kakigi, 1999; Pylkkänen, Stringfellow, & Marantz, 2002; Pylkkänen & Marantz, 2003; Pylkkänen, Feintuch, Hopkins, & Marantz, 2004; Pylkkänen, Llinás & Murphy, 2006; Stockall, Stringfellow, & Marantz, 2004; Sekiguchi, Koyama, & Kakigi, 2001; Simos, Breier, Fletcher, Foorman, Castillo, & Papanicolaou, 2002; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999, among others) identifies a series of components in the MEG waveform following visual word onset which map onto different subprocesses during the time course of visual word recognition. The first component, a bilateral occipitotemporal component around 150–200 ms post word onset, reflects pre-lexical properties of the visual word stimulus, such as letter-string length and letter position effects (Tarkiainen et al., 1999; Cornelissen et al., 2003, among others). A second component, peaking between 200–300 ms with a complex of underlying neural sources in the posterior portion of the left hemisphere, is less well understood but has

shown sensitivity to prelexical phonological properties of word stimuli (Pylkkänen et al., 2002). The third component, called the M350, is a response peaking approximately 350 ms post-onset of a stimulus, with a left superior temporal cortex generator. Studies have shown its sensitivity to factors such as word frequency, repetition priming, semantic relatedness, morphological relatedness, morphological family frequency and size, and properties of root semantics such as homonymy and polysemy (see Beretta, Fiorentino, & Poeppel, 2005; Pylkkänen et al., 2006, among others). As argued by Pylkkänen and Marantz (2003), this response may be conceived as a sub-component of the N400 response seen in ERP, a broad distribution appearing from around 200–600 ms post-onset of a visual word stimulus, which is thought to be sensitive to both semantic integration and automatic lexical access properties.

There are also studies which argue for very early responses sensitive to length and frequency (around 125–175 ms post-stimulus onset), whereas the literature cited above concludes that responses in this time window reflect aspects of visual word form processing. For example, Assadollahi and Pulvermüller (2003) conducted an MEG study contrasting responses to short and long words of high and low frequencies. Each condition contained 4 words, repeated multiple times. The authors report length effects after 100 ms post-stimulus onset, but also report frequency effects for the short words in this time window; effects for long words were detected later.[10] What subroutine of lexical access happens when in time, and is indexed by what dependent measure, covers a range of interesting research questions that remain under current investigation. As it is beyond the scope of this paper to engage in all such questions, we note that the present study aims to test one particular hypothesis about one aspect of morphologically complex word processing, namely whether contrasts of whole-word vs. morphemic constituent properties results in a divergence in response times and/or

---

[10] As regards the issue of when the *first* contact with the lexicon occurs, it is of the utmost concern to disentangle effects of visual word form properties with lexical properties such as word frequency. Given that the MEG responses from 100–200 ms are sensitive to properties of the visual word form (letter-length, discriminability in low-contrast presentations, etc.) it is important to consider the possible role of these properties when testing for early frequency effects. For example, while the length and frequency controls were reported in Assadollahi and Pulvermüller (2003), there is no mention of orthographic/phonological regularity or probability controls. On the view that these properties are likely to correlate with frequency, this is of concern given the many studies showing the components around 100–200 ms post-onset are sensitive to letter-string encoding (e.g., Cornelissen et al., 2003; Tarkiainen et al., 1999), making the interpretation of the effects as lexical-level more difficult. Further differences which are potentially relevant, as noted by Assadollahi and Pulvermüller (2003), are the differences in design among the studies showing earlier vs. later effects of frequency, and differences in how the frequency effects that were reported were reflected in the signal (amplitude modulation vs. latency modulation).

components which have been identified in the MEG signal preceding the behavioural response; we do not aim to test for, nor argue for, what is the absolute earliest point at which it is possible to elicit any putatively lexical effect in any electrophysiological study (see also Hinojosa, Martin-Loeches, Muñoz, Casado, & Pozo, 2004; Martín-Loeches, Hinojosa, Gómez-Jarabo, & Rubia, 1999; Sereno, Rayner, & Posner, 1998, for further discussion of early lexical effects in electrophysiology). Rather, in the current study, we focus on a narrow set of hypotheses on the processing of complex words, contrasting models which predict that morphemic properties should affect the components along the time course of lexical processing (accounts positing rapid and automatic decomposition) with those which do not ascribe a role for constituents in online lexical processing (accounts claiming non-decompositional processing during initial access to the lexicon).

In sum, including electrophysiological recordings in our experimental paradigm allows for the testing of each of the electrophysiological components along the way to a participant's lexical decision for subtle differences (if any) reflecting the stimulus manipulation (our whole-word vs. constituent contrast), with *directional* predictions on how such effects should look and pattern with the response time measure, under competing conceptions on the role of morphological decomposition (see Figure 3). The 350 ms MEG component is important in this regard, given its sensitivity to lexical properties, as reported in the literature summarised above, and its location in the time course (prior to the overt response time measure) for the investigation of effects of morphological structure in word recognition.
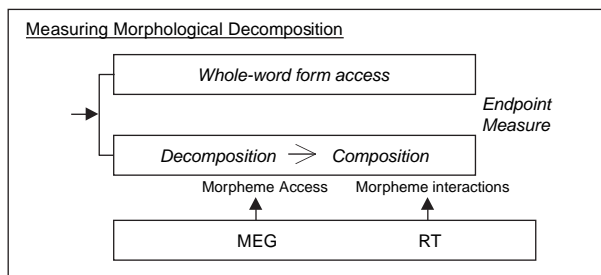
Figure 3. Response time is a single measure that may reflect both early and late processing; magnetoencephalographic (MEG) signals, which can be analysed throughout the time course from the onset of a word through the response, have the potential to reflect various aspects of processing, including those preceding the overt response.

THE PRESENT STUDY

The present study utilises simultaneous lexical decision and brain-level (MEG) measures to track the time course of decomposition in compound words. In this way, it is possible to measure access to constituents with the properties of specific components in the electrophysiological signal and their patterning with and divergence from the RT data, as predicted under competing conceptions of morphological decomposition.

We test for the effects of compound word structure by pairwise matching *single words* and *lexicalised compound words orthographically written as one word* on overall properties thought to affect access. Crucially, the morphemic constituents of the compounds are mismatched to the single words such that their morpheme-level properties would give them an advantage in access, under a decompositional approach, as shown in Figure 1. We show that morphological structure is reflected in the combined brain and behavioural measures in a manner that rules out non-decompositional theories and is most consistent with models incorporating early effects of abstract morphological structure.

Materials and methods

Participants. Twelve right-handed, monolingual American English-speaking adults with normal or corrected-to-normal vision (eight females; ages 18 to 26, mean age 21) provided their written informed consent to participate in this experiment. Participants were paid for their participation.

Materials and design. The materials were 120 word items, drawn from the Collins Cobuild English corpus (320 million words), and 120 non-word items, approximately one-half of which were formed from orthographic transcription of non-word items from Vitevitch and Luce (1999). The 120 word items comprised 60 disyllabic, single (monomorphemic) words and 60 bi-morphemic noun-noun compounds orthographically represented without

TABLE 1
Samples of visual lexical decision stimuli

Condition	Mean Log. Freq.*	Mean no. letters	Example
Compound (CW)	0.451	7.82	flagship
CW 1st / 2nd *constituents*	*1.96/1.98*	*3.82/4.0*	*flag/ship*
Single word (SW)	0.459	7.78	crescent
Nonword (NW)		7.81	nishpern
W-NW Foil (WNW)		7.94	crowskep

* Parts per million (ppm): CW 2.82 ppm; CW 1st / 2nd constituents 91.2/95.5 ppm; SW 2.88 ppm.

spacing as drawn from the Collins Cobuild corpus. See Appendix I for the complete list of items.

The compounds and single words were pairwise matched on frequency[11] (compound mean frequency = 0.451, single word mean frequency = 0.459, $t < 1$), overall letter-length (compound mean length = 7.82, single word mean length = 7.78, $t < 1$), and syllabicity (exactly matched in every case, all disyllabic).[12]

The compound words were selected such that the first- and second-position constituents had higher log frequency (first constituent mean log frequency = 1.96, second constituent mean log frequency = 1.98) compared with the overall compound word and matched single word frequency. This frequency manipulation resulted in a large frequency mismatch with both first constituent (C1) and second constituent (C2) having significantly higher frequency, as shown by Analysis of Variance (see Appendix II for statistical tests on all the item controls reported in this section).

The same is true of length in letters (first constituent mean letter length = 3.82, second constituent mean log letter length = 4.00) in comparison with the overall compounds and matched single words (compound letter length = 7.82, single word letter length = 7.78). The constituents were significantly shorter (see Appendix II). All constituents were monosyllabic and all whole compounds and single words were disyllabic.

Frequency and length-matched subsets of 12 pairs of compound and single words from among the top, middle and bottom 20% log-frequency were selected for subanalyses among compounds and single words within high, middle, and low overall frequency levels. (The terms high, middle, and low are meant with respect to the stimuli being categorised here, as a way to identify subsets at three relatively different frequency strata within the stimulus set.)

---

[11] The frequency counts used here are lemma frequency (the frequency of a word form and its inflectional variants). The lemma frequency counts for whole compounds and single words are counts for the whole word form uninterrupted e.g., by hyphenation, that is, in the same form that they are presented in the experiment.

[12] Note that Cobuild is advantageous due to its extremely large sample size, allowing a good estimate of word frequency among words not at the top of the frequency range, like the current items. However, we consulted the Francis and Kučera (1982) analysis of the Brown Corpus (first published 1961; approximately 1 million tokens) for the sake of comparison. Only 60% of the compound words, and 70% of the single words were represented in that corpus. Nevertheless, measures of frequency (log frequency, raw frequency) and distribution (number of sources in which lemma appears, number of text samples in which lemma appears) showed that the subsets of these items were matched; log frequency differed by only 0.09 across conditions among the extant CW and SW on this count. In the Carroll, Davies, and Richman (1971) American Heritage Intermediate Dictionary corpus (approximately 5 million tokens), approximately 85% of the stimuli were represented in the corpus. The mean log frequency of the extant CW and SW was again matched; log frequency differed by only 0.05 across conditions.

Mean letter-length for the compound vs. single word comparisons was matched identically within each of the three frequency subsets of the single and compound words, at 8.08 letters for high frequency, 7.75 letters for medium frequency, and 7.17 letters for low frequency (only the contrast among high- and low frequency items was significant in a planned comparison, as shown in Appendix II). As before, overall and constituent properties were mismatched, with constituents having significantly higher frequency, shorter length, and fewer syllables (see Appendix II for details).

The 120 non-words included 104 disyllabic pronounceable nonwords, approximately one-half of which were built using orthographic transcriptions of monosyllabic nonwords from Vitevitch and Luce (1999). Mean letter-length of these 104 nonwords was 7.81 (word vs. nonword length matched, $t < 1$). Sixteen additional non-words formed the category of Word-Nonword foils. These non-words contained a monosyllabic English noun in the first-syllable position and a monosyllabic nonword second syllable. Mean log frequency of the morphemic first syllable of the word-nonword foils was 1.48 ($SD = 0.84$), and mean letter-length was 3.81 ($SD = 0.66$). Overall mean length for the word-nonword foils was 7.93 ($SD = 0.68$).

Predictions

Response time.   Since RT is sensitive to early and to late lexical processes, either an early decomposition (decomposition-first or parallel dual-route model) or a late decomposition (decomposition-second) account allows for the prediction that compounds will differ from single words, due to the properties of the constituents rather than the whole word. A non-decompositional account would predict no differences due to word structure, as the overall word properties are matched.

Electrophysiological predictions.   The latency of the MEG component at 300–400 ms after word onset indexes processing related to lexical access rather than post-lexical processing (Embick et al., 2001; Pylkkänen et al., 2002). Given this measure, early decomposition predicts an effect not only in response time but also in the M350 component, reflecting constituent over whole-word properties. Late decomposition predicts response time differences but, like the non-decompositional account, no M350 divergence. Further, since the CW are lexicalised and have short constituents, lexicalisation and length constraints predict no RT or M350 differences.

Given the properties of our stimuli (mismatches among whole-word and morphemic constituent properties) and the two types of measurement used (response time and electrophysiology), we thus test the following directional predictions:

a. Compounds should overall be faster than single words in both RT and M350 latency, if decomposition is deployed online and early in time course.
b. Compounds should overall be faster than single words only in RT, if at all, if decomposition is deployed online but late in time course.
c. The response time advantage should not persist for those items for which early access to constituents may not facilitate response to an internally structured representation, such as the word-nonword foils.

Procedure and recording

Stimuli were visually presented using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993) in a randomised order, in three blocks of 80 stimuli with a pause after each block to provide resting time for the participant. The experimental paradigm was continuous lexical decision, for which the participants were instructed to decide whether each item was a word or a nonword. Each trial was initiated with a 1000 ms fixation point in the centre of the screen, followed by visual presentation of the stimulus, lasting until participant's response via button press. The intertrial interval was varied pseudorandomly among values at 50 ms intervals between 500 and 1000 ms. 'Word' responses were made by button-press using the dominant (right) hand, 'Nonword' responses by button-press with the non-dominant (left) hand.

During the experiment, the participants lay in a dimly lit magnetically shielded room, viewing items presented on a screen fixed 37 cm above the participant's eye-level. The text was presented in Geneva font, size 48, in Magenta letters on a black background. Words subtended approximately $1.4°$ vertically and $6.4°$ horizontally (range $4.6°–8.6°$). Button-press responses were made using a two-pad non-magnetic fibre-optic response-button system (Current Designs, Inc., Philadelphia, PA).

Neuromagnetic signals were recorded continuously with a 160-channel whole-head axial gradiometer MEG System (Kanazawa Institute of Technology, Kanazawa, Japan). Prior to the recording, five electromagnetic coils were positioned on the participant with respect to anatomical landmarks: the nasion, preauricular flaps, and two forehead positions. The nasion and pre-auricular points were then digitised, as was location of each of the five coils. The location of these coils with respect to the sensors was recorded immediately before and after the experimental recording for subsequent coregistration with digitised headshape or MRI images, to make possible analyses of these data which may involve source localisation.

Data were recorded at a sampling rate of 1000 Hz, filtered online with a band-pass filter of 1–200 Hz and a band-elimination filter at 60 Hz. The continuous data file was then noise-reduced relative to three reference

magnetometer coils using the Continuously Adjusted Least Squares Method (CALM; Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001). Trials were then averaged by condition with epochs beginning 100 ms before stimulus onset and extending to 600 ms post-onset. The trials were then level-rejected at $+/-2.5$ pT to remove trials with eye-blinks or other artifact, if any. The averaged data were baseline-corrected relative to the 100 ms prestimulus interval, and low-pass filtered at 30 Hz.

Results

Response time and accuracy

Response times and accuracy were analysed for each participant as follows. Data from incorrect trials (approximately 5.5% of data) and responses differing by more than 2 standard deviations from the condition mean (overall word category CW, SW, Word-Nonword Foil, Nonword) were removed from participants' response time results (approximately 3.7% of the data points). Response time was compared among the single words and compound words, and among the compound words and single words vs. word-nonword foils. The mean response times and accuracy rates are shown in Table 2.

Response time.   Response time differed significantly by word category, $F(2, 22) = 24.057$, $MSE = 5708.597$, $p < .001$. Planned comparisons showed that compounds ($M = 672$ ms) were responded to faster than single words ($M = 743$ ms), $F(1, 11) = 42.103$, $MSE = 722.39$, $p < .001$, that compounds were responded to more quickly than word-nonword foils ($M = 882$ ms), $F(1, 11) = 31.402$, $MSE = 8452.686$, $p < .001$, and that single words were responded to more quickly than word-nonword foils, $F(1, 11) = 14.608$, $MSE = 7950.643$, $p < .004$. Non-foil nonword fillers were responded to with a mean RT of 793 ms ($SE = 35.6$).

Accuracy.   A direct comparison of accuracy among compounds, single words, and word-nonword foils is telling about differences due to the

TABLE 2
Response times (mean, *SE*) and accuracy (Pct.) for compounds, single words, word-nonword foils, and other nonwords

Word category	Mean RT (ms)	SE	Accuracy (%)
Compound	672	29	97%
Single word	743	36	90%
Word-nonword foil	882	47	99%
Other nonwords	793	36	99%

TABLE 3
Response times (mean, *SE*) and accuracy (Pct.) for subsets of the compounds and single words at three frequency levels

Word category	Mean RT (ms)	SE	Accuracy (%)
CWH	624	23	99%
SWH	678	30	100%
CWM	670	32	99%
SWM	749	39	90%
CWL	731	36	94%
SWL	774	42	80%

representation of word structure, and also as to whether foil accuracy shows a speed-accuracy tradeoff or strategic response strategy. The omnibus ANOVA for accuracy shows a significant difference by word category, $F(2, 22) = 13.010$, $MSE = 0.002$, $p < .001$. Planned comparisons showed that the compound words ($M = 97\%$) were responded to with higher accuracy than single words ($M = 90\%$), $F(1, 11) = 13.933$, $MSE = 0.002$, $p < .004$. Compounds and foils ($M = 99\%$) did not differ significantly in accuracy, $F(1, 11) = 3.887$, $MSE = 0.001$, $p < .075$, although single words were responded to with lower accuracy than word-nonword foils, $F(1, 11) = 14.668$, $MSE = 0.003$, $p < .004$.

Non-foil nonword fillers were responded to with a mean accuracy of 99% ($SE = 0.5\%$). Note that the compounds and word-nonword foils were both responded to with high accuracy, suggesting that a strategy based on word-like first syllable is not at play.[13] Figure 4 shows the response time and accuracy rates for compound words, single words and word-nonword foils.

Groups of 12 words each from among the highest, middle and lowest 20% of pair-wise log-frequency matched single words were selected for separate response time and accuracy analysis. The mean response times and accuracy rates are reported in Table 3.

---

[13] While the fact that nonword responses were made with the left (non-dominant) hand makes the interpretation of the response time slowdown more complicated since it requires comparing responses across hands (however see Taft & Forster, 1976 among many others for the same result), the high accuracy on this condition provides support to the notion that compound responses were not entirely driven by spotting a morpheme. As for the crucial compounds vs. single words comparison, both responses were made on the dominant (right) hand. One way to address concerns about response hand would be to vary response hand by participant or block; this solution was not utilised in the current study in order to avoid across-participants or across-block analyses of the electrophysiological data (MEG), as across-participants comparisons are non-standard, and number of carefully matched samples should be maximised to achieve the highest signal-to-noise ratio in the MEG responses.
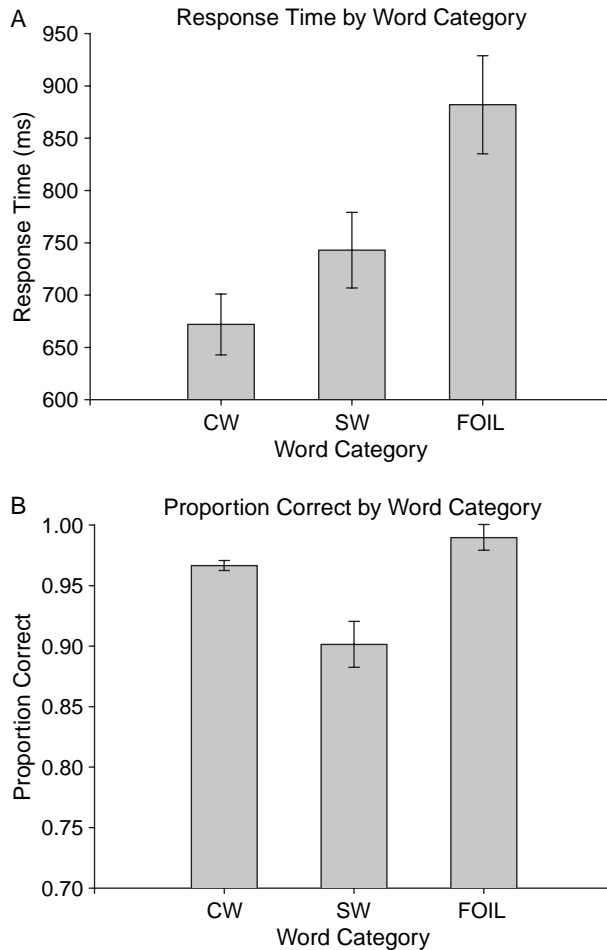
Figure 4. Response time (Fig. 4A) shows a significant difference by word structure, with compounds responded to faster than both single words and word-nonword foils. Accuracy (Fig. 4B) was very high for the compounds and for the word-nonword foils. Single words were responded to with lower accuracy overall.

Response time (subanalyses at three frequency levels). The overall ANOVA (2 word structures × 3 frequency levels) revealed significant effects of word structure, $F(1, 11) = 27.979$, $MSE = 2218.071$, $p < .001$, and of frequency level, $F(2, 22) = 16.863$, $MSE = 3721.713$, $p < .001$, although the structure × frequency interaction was not significant, $F(2, 22) = 1.355$, $MSE = 1598.517$, $p < .280$. The effect of frequency was significant in planned comparisons among high vs. mid frequency, $F(1, 11) = 13.208$, $MSE = 3133.386$, $p < .001$, with a non-significant interaction with word structure,

$F(1, 11) = 1.163$, $MSE = 7239.796$, $p < .305$, and significant in a planned comparison of mid vs. low frequency, $F(1, 11) = 8.786$, $MSE = 2543.013$, $p < .014$, again with a non-significant interaction with word structure, $F(1, 11) = 3.050$, $MSE = 5338.474$, $p < .11$. Planned comparisons among compounds and single words at each frequency level show significant RT facilitation for high frequency compounds ($M = 624$) relative to single words ($M = 678$), $F(1, 11) = 11.943$, $MSE = 1429.44$, $p < .006$), significant RT facilitation for middle frequency compounds ($M = 670$) relative to single words ($M = 749$), $F(1, 11) = 28.431$, $MSE = 1344.702$, $p < .001$, and marginal RT facilitation for low frequency compounds ($M = 731$) relative to single words ($M = 774$), $F(1, 11) = 4.198$, $MSE = 2640.962$, $p < .066$.

Accuracy (subanalyses at three frequency levels). The ANOVA for accuracy among the three frequency subsets (2 word structures × 3 frequency levels) revealed significant effects of word structure, $F(1, 11) = 13.644$, $MSE = 0.008$, $p < .005$), of frequency level, $F(2, 22) = 15.293$, $MSE = 0.007$, $p < .001$, and a significant structure × frequency interaction, $F(2, 22) = 11.396$, $MSE = 0.003$, $p < .001$. The effect of accuracy was significant in planned comparisons among high vs. mid frequency words, $F(1, 11) = 18.526$, $MSE = 0.002$, $p < .002$, with a significant interaction by word structure, $F(1, 11) = 13.146$, $MSE = 0.009$, $p < .005$, and for mid vs. low frequency words, $F(1, 11) = 8.741$, $MSE = 0.007$, $p < .014$, with a non-significant interaction, $F(1, 11) = 3.667$, $MSE = 0.013$, $p < .083$. Planned comparisons among compounds and single words at each frequency level show a non-significant accuracy difference for high frequency compounds ($M = 99\%$) relative to single words ($M = 100\%$), $F(1, 11) = 1.000$, $MSE = 0.000$, $p < .340$, but a significant effect for middle frequency compounds ($M = 99\%$) relative to single words ($M = 90\%$), $F(1, 11) = 10.385$, $MSE = 0.005$, $p < .009$, and for low frequency compounds ($M = 90\%$) relative to single words ($M = 80\%$), $F(1, 11) = 14.011$, $MSE = 0.01$, $p < .004$.[14] Comparisons among compound words and single words at three frequency levels with respect to response time and accuracy are shown in Figure 5.

The behavioural results, in summary, support the decompositional view of morphological processing. The response times to compound words were faster, reflecting the influence of morphemic constituent properties rather than only whole-word properties. The fact that the response time facilitation for compounds persisted even among the highest overall frequency

---

[14] We also report the results of a replication study with 12 additional participants, in Appendix III, Tables 5A–C. The pattern of results is identical to that of the current experiment, both in a by-participants and in a by-items analysis. Further, these data were reanalysed excluding six single-word items which might be analysed as complex (opaque/bound forms); the patterns were the same. These data are reported in Appendix III, Table 5C.

A

Response Time by Word Category
at Three Frequency Levels

B

Proportion Correct by Word Category
at Three Frequency Levels
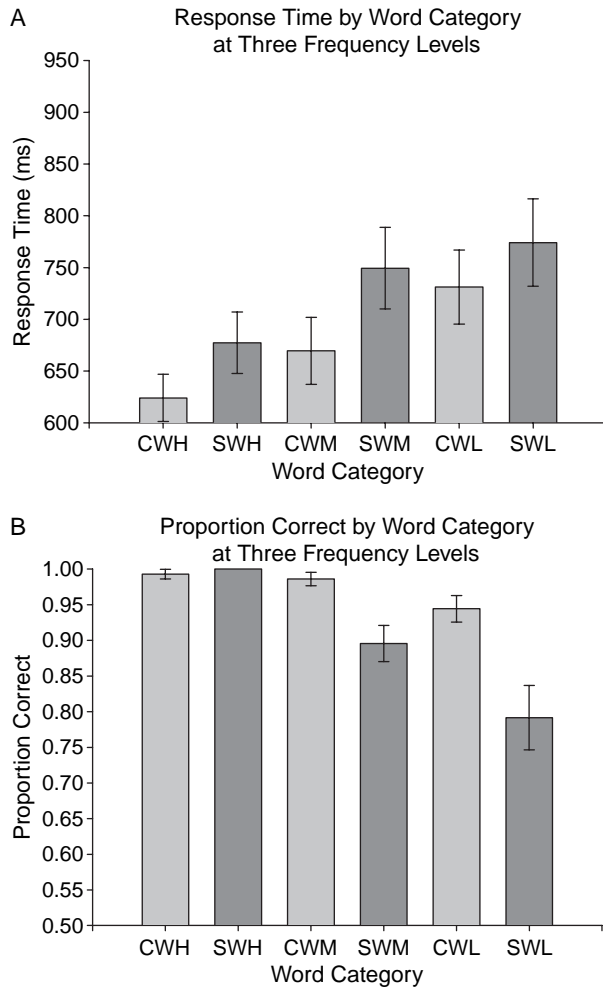
Figure 5. Response time (Fig. 5A) shows a significant difference by word structure, with compounds responded to faster than single words across the three frequency levels. (RT was not significantly different, but marginal, among the lowest frequency CW-SW.) Figure 5B shows accuracy for subsets of the items at three frequency levels, showing that accuracy diverges among the middle and low frequency subsets, suggesting that compounds decrease in accuracy with a shallower slope than single words.

compounds is particularly challenging to full-storage models like Bybee (1995) that predict more whole-word and less decomposition-like processing as surface frequency increases, as whole-word representations will be strengthened and relations with morphemes will be weakened. They will also be challenging for versions of the dual-route model which take surface

frequency as a determiner of decomposability (for example Stemberger and MacWhinney, 1986, among others). Moreover, compound words were accurately judged as 'word' stimuli at significantly higher rates, suggesting a qualitative difference in the items by word structure. The findings support the view of morphologically complex items processed as internally structured representations in the mental lexicon. Access to morphemic constituents facilitated response time and resulted in higher accuracy for compounds relative to overall matched single words. Word-nonword foils, however, resulted in delayed RT but high accuracy, as has been reported previously for this type of item (and for word-word novel compounds). This suggests a role for decomposition even in the parsing of known compounds, as proposed by decomposition first/full-parsing accounts.
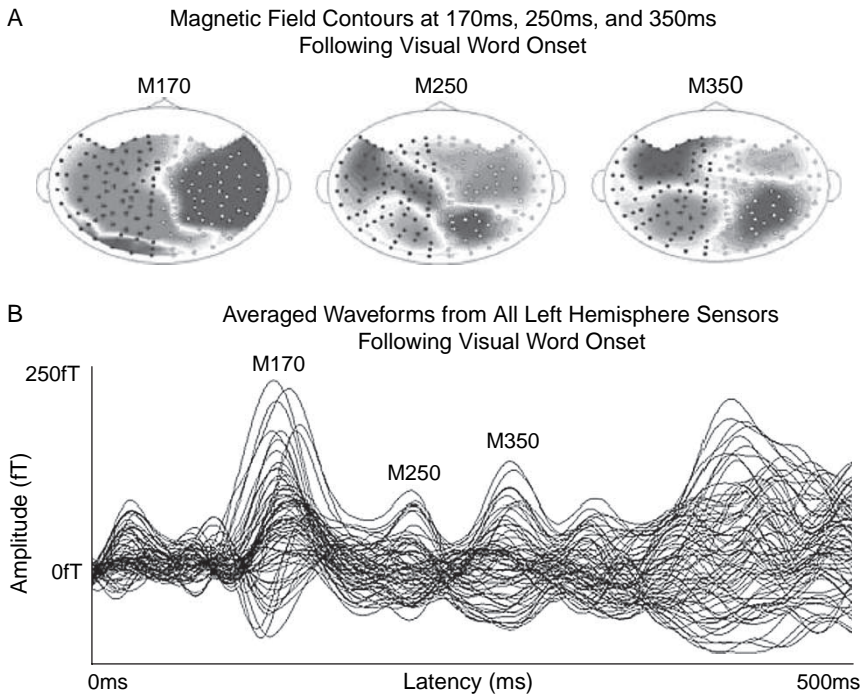
As argued above, it is difficult to place this effect as early or late in time course using lexical decision data alone, although the pattern of effects strongly suggests some version of an early decompositional parser, and it would be difficult for a late, decomposition-second parser to capture these effects. If the electrophysiological signal shows a difference favouring the compound constituents in a component in the evoked waveform thought to underlie lexical access, however, the results may be more conclusive in supporting an early decompositional version of the decompositional parser. Undoubtedly, MEG does not index only early processes. However, it has been hypothesised that one or more parts of the electrophysiological signal may yield an index of constituent access, capturing a subcomponent of decompositional processing that may not always be detectable at lexical decision. In the following, we analyse the electrophysiological signals leading to the lexical decision response with respect to both constituent and whole-word properties.

Neuromagnetic signals

Analysis of the magnetoencephalographic signal revealed three consistent components in cascade from the onset of the visual stimulus through the first 500 ms post-onset, as shown in Figure 6. The occurrence of this series of components, appearing at approximately 170, 250, and 350 ms post-onset was consistent across conditions and participants, and is also attested in a growing cohort of neuromagnetic studies (Embick et al., 2001; Beretta et al., 2005, among others).

The peak latency and amplitude for each component was determined by selecting five channels from the sink (ingoing) and five channels from the source (outgoing) portion of the magnetic field contour; the latency of the peak from a root mean square (RMS) analysis on these ten channels was entered into by-condition statistical comparisons.

A

Magnetic Field Contours at 170ms, 250ms, and 350ms Following Visual Word Onset

M170   M250   M350

B

### Averaged Waveforms from All Left Hemisphere Sensors Following Visual Word Onset

Figure 6. Three components in the magnetoencephalographic (MEG) response to visually presented word stimuli are typically observed. (A) Contour maps at response peak (the light grey areas represent the outgoing portions of the magnetic field contour, and the dark grey areas represent the ingoing portions of the magnetic field contour). The first distribution, around 170 ms post-onset, is associated with visual word form processing. The response has been argued from dipole modelling to originate in the occipito-temporal cortex, perhaps in the fusiform gyrus. The component around 250 ms post-onset may reflect some aspects of phonological or ortho-phonological processing. The third component, peaking between 300–400 ms has been previously implicated as involved in lexical access. This was the first component sensitive to the stimulus manipulation in the current study. (B) Butterfly plot of all channels over left hemisphere illustrating peaks in the evoked field.

Electrophysiological results: M350. The third source component in the pattern of distributions was the first component sensitive to the compound versus single word comparison. This component, peaking around 350 ms, yielded a significant effect of condition in ANOVA, $F(2, 22) = 8.532$, $MSE = 288.114$, $p < .003$. Planned comparisons show a significantly earlier peak latency for the compound words ($M = 333$ ms) than the single words ($M = 360$ ms), $F(1, 11) = 12.732$, $MSE = 354.223$, $p < .005$. Word-nonword foils ($M = 340$ ms) did not differ significantly from compound words, $F(1, 11) = 1.394$, $MSE = 186.587$, $p < .264$. Single word latency was significantly longer than word-nonword foil latency, $F(1, 11) = 8.049$, $MSE = 323.53$, $p < .017$.

A   Magnetic Field Contours at RMS Peak for CW (left), SW (center)
and Word-Nonword Foils (right)

B   RMS Waveforms for CW vs. SW (top)
and Word-Nonword Foils vs. SW (bottom)
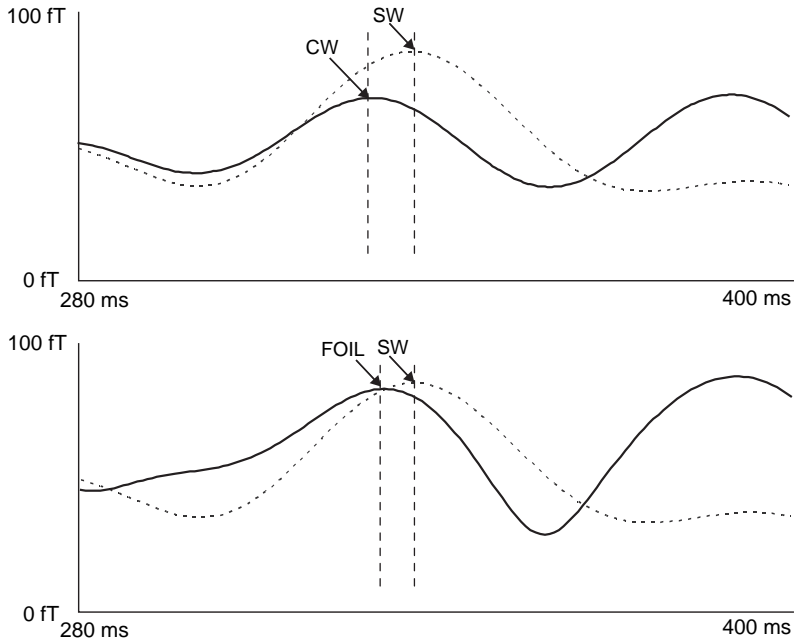
Figure 7. The magnetic field contours for compounds, single words, and word-nonword foils are plotted in Fig 7A. (the light grey areas represent the outgoing portions of the magnetic field contour, and the dark grey areas represent the ingoing portions of the magnetic field contour). The whole-head images show very similar magnetic field contours underlying the RMS waveforms for the responses across the three conditions. The root mean square (RMS) waveforms, representing an averaged signal from 10 sensors, are plotted individually for compounds vs. single words in 7B (top), and for word-nonword foils vs. single words in 7B (bottom). While other variations in the RMS waveform did not consistently differ by condition across participants, the latency of the RMS peak in this time window is consistently earlier for the compounds than for the single words, and earlier for word-nonword foils than single words.

TABLE 4
Mean peak latency and SE, for compounds, single words and word-nonword foils

Word category	Mean latency (ms)	SE
Compound	333	13
Single word	361	16
Word-nonword foils	340	12

Table 4 shows the mean differences in M350 latency, and the MEG waveforms are shown in Figure 7.

Electrophysiological results: Earlier components (M170, M250). Analysis of the first two components appearing consistently across conditions and participants, peaking around 170 ms post-onset (M170), and 250 ms post-onset (M250), were tested using the same method as that for the M350 response. Neither the M170 nor the M250 yielded significant latency differences by condition (all $Fs < 1$). Likewise, peak amplitude did not differ at M170, M250, nor did it differ at M350 (all $Fs < 1$). The direction of M350 latency and response time differences across conditions are depicted in Figure 8.

Summary and limitations of the electrophysiological results. Three components in the averaged evoked MEG waveform were observed consistently across conditions and participants, one peaking around 170 ms (M170), one peaking around 250 ms (M250), and a third peaking around 350 ms post-onset of the visually presented words. Of these, only the third component showed a significant effect of word structure in the RMS analysis, the M350 component. While these results support a particular timing prediction, namely that M350 peak latency should show an effect of facilitation due to the properties of compounds' morphemic constituents when contrasted with disyllabic single words, these results leave open the precise nature of the underlying sources for the M350 and the other components consistently observed in the dataset.

Several studies have specifically addressed the underlying sources of these components. Tarkiainen et al. (1999), Helenius et al. (1999), and others have explored the localisation of the responses before 200 ms in response to visual words and symbol strings; from these studies, the activation around 170 ms has been attributed to inferior occipito-temporal cortex. Helenius et al. (1999), among others, have also explored the source localisation of the magnetic N400. Typically, this component has been localised broadly to left superior temporal locations, but with large individual differences (see e.g., Helenius et al., 1999 for one example, and Van Petten & Luka, 2006 for a
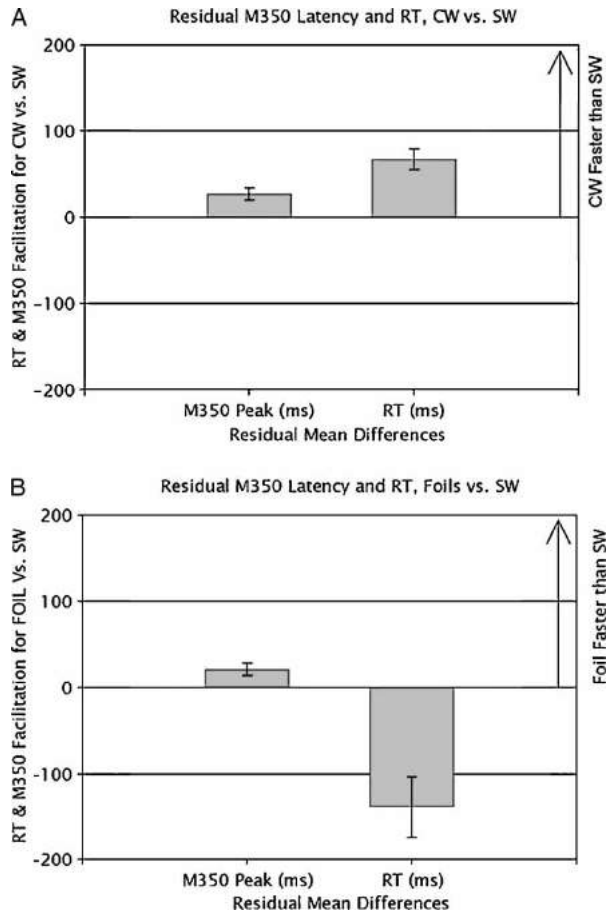
Figure 8. Reaction times are faster for compounds than single words. However, reaction times are slower for word-nonword foils than for compounds or single words. In the magnetoencephalographic (MEG) response, compounds are also faster than single words; crucially, in this measure, word-nonword foils are also faster than single words, indicating this component's sensitivity to a specific subcomponent of lexical processing reflecting morpheme-level activation.

recent review of source localisation findings regarding the N400m response). For example, in Helenius et al. (1999), a number of sources contributing to the N400-like response to semantically anomalous sentence-ending words were clustered around left superior temporal regions (anterior, middle and/or posterior), but context sensitive responses also appeared in many different regions – for two subjects in left frontal regions, for five subjects in regions posterior to the Sylvian fissure, and for five subjects right-side STG activation (localised with 100 trials/condition). Halgren et al. (2002) also

characterised the underlying sources of the responses to visual words over time in MEG using a variant of SPM (Statistical Parametric Mapping), again suggesting that the responses to visual words after 200 ms are quite diffuse in source localisation in the left hemisphere and bilaterally; Marinkovic, Dhond, Dale, Glessner, Carr, and Halgren (2003) also report distributed activation after 200 ms throughout anterior through posterior superior temporal regions, as well as inferior and medial prefrontal activation, as well as right hemisphere prefrontal and superior temporal activation. This has also been the case in previous attempts to localise the M350, in particular, as well as the M250 (see, e.g., Pylkkänen, Llinás, & Murphy, 2006, among others), both of which showed diffuse and variable source localisation across participants, including areas around left and right superior temporal regions, regions posterior to the Sylvian fissure, occipital regions, and left frontal regions (localisations from grand-average of 252 trials).

An exploratory source analysis of our data (using single equivalent current dipoles) suggests that, consistent with the previous studies, the M350 source tends to localise to left temporal regions, but with large individual variability. We do not pursue the analysis of the underlying sources of these components further, considering the following limiting factors for the current dataset in this respect. In our experience, given that our study required careful stimulus control to allow for testing of our psycholinguistic hypothesis (see e.g., Assadollahi & Pulvermüller, 2003 for more on this concern), this is not a dataset that is well-suited for a detailed source analysis of these complex components; the experiment was not designed to test for this, but to test a hypothesis about the timing of the peak activation of the M350 component in the evoked waveform. As previous studies have shown that they involve massive individual differences in source localisation, it is clear that further MEG experiments testing the underlying sources of these components are necessary. Indeed, MEG is particularly well-suited among electrophysiological methods, for undertaking this, but it will require higher numbers of trials and different stimuli, in order to obtain the optimal signal-to-noise ratio to permit believable source reconstruction for the components whose timing behaviour was measured in the current study. While precise source reconstruction is beyond the scope of the current paper, further research is needed on the sources underlying these components, both to increase our basic understanding of these components not only in time but in space, and also to yield testable, principled hypotheses about how the underlying source distributions may or may not contrast in interesting ways according to psycholinguistic variables and the putative functional roles of the underlying source components (see, for example, Hauk, Johnsrude, & Pulvermüller, 2004 for one recent study which tests the relation among motor cortex activation and the distinction among verbs in terms of their motoric decomposition).

Discussion

The response time and magnetoencephalographic results from this study favour a decompositional account of lexical processing and a model of word recognition incorporating early decomposition into morphemic constituents. The visual lexical decision results show (i) a response time advantage for the compound words over matched single words and (ii) accurate though delayed responses to pseudomorphemic nonwords. The former provides new evidence for morpheme-based lexical processing, and the latter reinforces the conclusion that the effect of morphological structure for compound response time is not just word finding in words.[15] The decompositional effects found in the current study support one fundamental aspect of the *full-parsing* approach, which is decomposition even for *lexicalised* words. The direction of the effect (facilitation) can be handled by positing the storage of internally structured entries which can be activated by prior access to constituents, and it is also consistent with a version of full-parsing without stored complex entries, namely one in which morpheme combination is not costly in the lexical decision task. With the behavioural measures, thus, we can argue for a qualitative difference among compounds, single words, and word-nonword foils motivated by their distinct processing profiles. These data provide further evidence for the decomposition of morphologically complex forms, using a method directly comparing complex and simplex words matched on overall properties, but contrasting in morphemic constituent properties. Our results are broadly consistent with the morphemic constituency effects reported in several studies, including Pollatsek and Hyönä (2005), Andrews et al. (2004), Juhasz et al. (2003), Shoolman and Andrews (2003), Zwitserlood (1994), and Andrews (1986), among others. In the following sections, we consider the implications of these results within the broader morphological processing literature.

The electrophysiological dependent measure (M350 latency) was hypothesised to track lexical access, with shorter, higher-frequency items expected to show earlier activation via peak latency. This allowed us to locate the effect of decomposition in time, since the same compound word suggests very different predictions for M350 latency in terms of its morphemic constituent properties and its whole-word properties. The compound words were significantly earlier in peak latency for this component than the single words, as predicted by the constituent properties rather than solely by the whole word properties; the MEG response patterns for word-nonword

---

[15] Compound nonwords with morphemic constituents consistently elicit long response times (Taft & Forster 1976, among others). Van Jaarsveld and Rattink (1988) and others have also shown that a novel compound's lexical status and its interpretability affect response times (see also Coolen, van Jaarsveld, & Schreuder, 1992, among others).

foils looked more like that of compound words than single words. The results suggest that this component reflects the specific aspect of the computation of compound structure which involves activation of morpheme-level constituents.

Morphological parsing

Finding a facilitative effect of internal structure in processing is consistent with the notion from recent studies that an early morphological parser is active in word recognition (Feldman, 2000; Feldman & Soltano, 1999; Frost et al., 1997; Frost, Deutsch, & Forster, 2000a; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Frost, Deutsch, Gilboa, Tannenbaum, & Marslen-Wilson, 2000b; Longtin et al., 2003; Marslen-Wilson et al., 1994; Rastle & Davis, 2003; cf. Plaut & Gonnerman, 2000). Rastle and Davis (2003) show this result in English in a masked priming paradigm using a short (52 ms) stimulus onset asynchrony (SOA). In masked priming, 'brother' primes 'broth', but 'brothel' does not prime 'broth'. Longtin et al. (2003) tested items with superficial morphological complexity in French and found priming whenever there is at least a surface string containing a legal root, even when the complex structure was only apparent and not accurate: *baguette* (gloss: little stick) is monomorphemic, but primes *bague* (gloss: ring). In contrast, there was no facilitation for words sharing only orthographic overlap without the apparent possibility of an exhaustive morphological parse (e.g., *abricot – abri*; gloss: apricot – shelter; -*cot* is not a suffix in French). In both the English and the French studies, the priming effect for morphologically complex and apparently complex words is seen both for semantically transparent morphologically structured primes (*departure – depart*) and semantically opaque primes (*department – depart*); in contrast, cross-modal priming tasks show that this priming only persists in cross-modal tasks for transparent items, at least in English and French (e.g., Marslen-Wilson et al., 1994; Longtin et al., 2003, Experiment II). These results together argue for an early structural morphological segmentation system. A similar pattern emerges for stem homograph priming. Badecker and Allen (2002) show that masked priming for stem homographs results in facilitation for the target. This effect holds for stem homographs: priming for *cerrar – cerro*; gloss: to close – hill, and is also dissociable from effects of semantic overlap (*puerta – cerro*; gloss: door – to close), while there is no significant facilitation for pairs with only orthographic overlap: *cerdo – cerro* (gloss: pig – hill). However, in longer-lag overt priming tasks, the previously facilitative stem homograph effect becomes one of inhibition (Allen & Badecker, 1999). Do the findings from this line of research in fact generalise to compounds? As noted above, recent masked-priming studies with compounds suggest that this effect indeed holds for compounds (Fiorentino,

2006; Shoolman & Andrews, 2003), further supporting the idea of an early morphological parse.

Other evidence comes from a recent production study (Roelofs & Baayen, 2002) which shows a preparation effect in the production of morphologically complex words, both semantically transparent ('input') and semantically opaque ('invoice'), but not monomorphemic words ('insect') (see also Levelt, 2001; cf. Jannsen et al., 2006). Additional support for a morphological but non-semantic effect in production comes from a study in Italian by Burani, Dovetto, Spuntarelli, and Thornton (1999), who show that pseudowords with morphological constituency are named more easily than pseudowords without morphological constituency.

The emerging picture suggests that an early but smart morphological parser is operative (as shown in studies of naming, masked priming, overt constituent repetition priming, fixation times in eye tracking, latency of the 350 ms MEG component), with initial parsing regardless of semantic transparency; effects of transparency begin to emerge, if at all, in measures which can also reflect subsequent stages of processing (such as cross-modal priming, semantic priming contrasting transparent and partially transparent vs. opaque compound primes, gaze duration in eye tracking, lexical decision).

Constraints

In addition to testing decompositional versus non-decompositional approaches, this experiment was able to test two putative constraints on the morphological influence in compound processing. The first is word length. Some previous studies (e.g., Bertram & Hyönä, 2003) suggested that word or constituent length may modulate morphological versus whole-word processing of compounds, and that the morphological effects seen for longer but not shorter words may be a non-structural effect arising from visual acuity (e.g., likelihood of needing two fixations versus one). The stimuli in our experiment were 'short' by the standard in Bertram & Hyönä (2003). Nevertheless, the effects of early access to constituents in the current study were as predicted under an early decomposition model, both in the electrophysiological and response time measurements. One advantage of the visual lexical decision/MEG method is that the starting points of processing and measurement (onset of visual stimulus) are clear and synchronised. In the eye-tracking methodology, both context and parafoveal preview likely play a role in affecting looking times; as regards parafoveal preview, the amount of information in the parafoveal view differs in size by condition in that study by virtue of the stimulus manipulation. While it may be complicated to directly compare the results among the two studies, the results together suggest some decompositional processing, and the results of

the current study suggest that, at least in lexical decision, even shorter compound words undergo early morphological decomposition.

The second constraint explored was lexicalisation. Contradictory effects in previous studies have given rise both to lexicalisation-invariant and novel-compound-only positions on decomposition (e.g., Van Jaarsveld & Rattink, 1988). However, the effects reported in the current study were all elicited with lexicalised compounds orthographically written as single words. The current study thus suggests that lexicalised compounds undergo early decomposition. As noted above, some models seek to capture putative morphemic constituency effects as effects of lexical relatedness, and predict that these relationships should be weaker for whole-word forms as their whole-word frequency increases (e.g., Bybee, 1995). Such models would have particular difficulty with the morphemic constituency effects observed for lexicalised compounds at the highest frequency level (e.g., *rainbow*, *baseball*) in the current study.

Naturally, more studies are necessary to further investigate the effects reported here. Some of the major outstanding issues are briefly discussed in turn.

Which constituent(s) drive the effect?.   In the current study, both the first and second constituents of the compounds were of higher frequency, shorter length, and fewer syllables than the whole-words. One could independently manipulate constituent properties by position to locate the effects for compounds as arising from first, second, or both constituent positions; the current data are agnostic on this point. However, there are constituency effects reported for both constituents in previous experiments manipulating frequency of compound constituents. For example, Juhasz et al. (2003) found effects for both first and second constituents of lexicalised English compounds in lexical decision, naming, and eye-tracking experiments, including particularly robust second constituent effects (see also Jarema, Busson, & Nikolova, 1999; Andrews et al., 2004; Pollatsek et al., 2000; Hyönä & Pollatsek, 1998; Pollatsek & Hyönä, 2005).[16]

Which properties of the constituents might drive the effect?.   Three factors were used to bias the constituents over the whole words: length, frequency, and syllabicity. Syllabicity has been shown to facilitate naming for difficult mono- and poly-syllabic words (e.g., Henderson, 1982; Taft, 1991). Length effects have been attested in both behavioural (Gill & McKeever, 1974; Lavidor & Ellis, 2002, among many others) and neurolinguistic studies

---

[16] Juhasz et al. (2003) presume that the relative pervasiveness of second position effects in their studies is because the second constituent position is where constituent and whole-word meanings converge in English (i.e., it is the head position).

(Cornelissen et al., 2003, Tarkiainen et al., 1999, among others). The MEG literature, for example, typically shows detection of word-length effects around 150–200 ms post-onset of visual word stimuli. However, frequency is the factor most often assumed to drive the effect in paradigms like the one used in the current study. Indeed, it has been common in the literature on lexical processing to contrast base and surface frequency of complex words (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bertram, Laine, Baayen, Schreuder, & Hyönä, 2000; Taft, 1979, among many others). In the literature on compounds, several studies have manipulated constituent frequency of compound words (Andrews, 1986; Juhasz et al., 2003; Pollatsek & Hyönä, 2005, among others). Many of these studies report base frequency effects on response times; however, accounting for the lack of base frequency effects under some circumstances has led to various dual-route models or full-storage models in some cases, although others object to the conclusion that lack of base-frequency entails a non-decompositional processing route (for more discussion, see for example Baayen, Dijkstra, & Schreuder, 1997; Bertram, Hyönä, & Laine, 2000; New et al., 2004; Taft, 2004).

Is there an independent index of composition as opposed to decomposition?. It would be valuable to identify an independent post-access electrophysiological index of composition in the MEG signal for stored entries and novel items. There have been some studies which speak to the role of composition for novel items ('parse time' in MRM; Schreuder & Baayen, 1995; for some linguistic analyses, see Downing, 1977; Levi, 1978; for some experimental results, see Gagné, 2002; Van Jaarsveld & Rattink, 1988, among others). From these results, we would expect novel combinations to be costly in some cases, as seems to be true for novel compounds, whereas initial access to constituents may be fast. In other cases, in contrast, access to morphemes would be facilitative, such as for known complex words. Behaviourally, facilitation would be more likely to be reflected in decision measures such as response time (RT) for known words, whereas in the novel case, true combination should be reflected in long RTs. Shoolman and Andrews (2003) suggest that 'combination' effects of the latter type can also be detected by increasing the proportion of nonwords with lexical routes, as suggested by the attenuation of base-frequency facilitation and priming for RT (Shoolman & Andrews, 2003).

We do not report results on an electrophysiological component indexing composition. However, two possibilities from the MEG literature are worth briefly speculating on here. The first is the possibility that a subsequent iteration of an M350-like distribution in the 400 ms range may be involved in this kind of process (Fiorentino & Poeppel, 2003; Ozawa, Fiorentino, & Poeppel, 2003; Pylkkänen & Marantz, 2003). The second, intriguing possibility is that the index of this kind of combinatorics lies in the

higher-frequency brain response (such as the gamma response, in the 20–50 Hz range). This response has been linked to binding in cognitive tasks in other domains, such as the visual domain (e.g., Tallon-Baudry & Bertrand, 1999). Some recent studies using EEG and MEG have begun exploring possible indices of linguistic properties in the gamma band (Braeutigam, Bailey, & Swithenby, 2001; Eulitz, Eulitz, Maess, Cohen, Pantev, & Elbert, 2000; Pulvermüller, 2002). If complex words are complex lexical items with internal structure computed in real time, these stimuli may be of use in testing general properties of decomposition and composition in brain-level computation (Gennari & Poeppel, 2003; Poeppel & Embick, 2005).

Processing models

The current processing models prominently featuring a decompositional component, the MRM, AAM, and the supralexical model of Giraudo and Grainger (2000) face problems accounting for the range of data in the literature in a principled way. A decomposition-second model would have trouble accounting for the data presented here and elsewhere without further stipulations. The data would be consistent with a dual route model such as MRM, since the data here support a class of models incorporating minimally a parallel decompositional component. As a model, MRM has the advantage of parallel availability of both options, but in order to accommodate the data in the literature, must direct traffic via constraints; further, at least some variants of the model do not predict decomposition for the items tested (lexicalised compounds). The effect of word structure, even among the highest-frequency compounds, supports full-parsing models (e.g., Taft, 1991, 2004; Stockall & Marantz, 2006); however, the facilitative effect relative to the single words requires a principled account under full-parsing. The findings of the current study are also compatible with some parallel dual-route or segmentation-through-recognition models which posit a stored representation with internal morphological structure which can be accessed via initial activation of morphemic constituents.

Semantic transparency

The focus of the present study was to test for the presence of decomposition in visually presented, lexicalised English compounds. Our behavioural and neural results suggest a role for early decomposition in the recognition of known compound words. While we did not test specifically the property of semantic transparency/opacity, it would be potentially informative to explore the role of transparency in compound representation and processing under similar conditions to those of the current study. For example, having both response-time and brain-level measures for potential effects of semantic opacity allows for further testing of the hypothesis that the relation among

the compound and its constituents is separable from semantic/phonological relations among the compound and its constituents, as suggested by masked priming effects regardless of transparency on the one hand, and transparency effects in cross-modal and overt semantic priming, on the other hand. The loci of these effects may be differentiable by measuring not only response time but also neural components involved in lexical activation, which may be informative regarding how we should incorporate such constraints into the parsing model.[17]

Visual vs. auditory processing

For auditory compounds, when we look at the speech stream in a phonemic way, we might expect that at some point a compound like 'teacup' is really only its first constituent: 'tea', since it is unfolding in time. If we were only mapping the constituents of the compound phonemically, then we might expect every auditory compound to be processed sequentially as a two-word sequence. However, if there are acoustic cues on the nonhead to signal otherwise, then we might expect an earlier commitment to compound interpretation than say the non-head's word offset. The latter kind of processing has been shown for German (Isel, Gunter, & Friederici, 2003, among others). Thus, the auditory processing of compounds provides converging evidence for the notion of immediate access to compound structure in processing.

Conclusion

The addition of simultaneous MEG recording to the lexical decision task offers a new way to investigate a deep property of lexical representation and processing, morphological decomposition, and points to a method for testing and constraining time course predictions on the role of decomposition in lexical processing. The behavioural and electrophysiological results support the growing variety of studies that suggest a role of

---

[17] Further, showing constituency effects regardless of transparency would seem to be at odds with the supralexical model, which posits constituent access after initial contact with whole-word representations and only for transparent words; such findings would also present challenges to distributed-connectionist approaches to constituency effects, since those models capture such effects as form-meaning overlaps. Elsewhere we report response time facilitation in masked priming both for constituents of transparent and of opaque lexicalised compound primes (Fiorentino, 2006); for further compound-constituency effects independent of semantic transparency, see also Shoolman and Andrews (2003) (masked priming from constituents to compounds), Zwitserlood (1994) (overt visual priming from compounds to constituents), Libben, Gibson, Yoon, and Sandra (2003) (overt visual priming from constituents to whole compounds) and Pollatsek and Hyönä (2005) (morphemic frequency effects on fixation durations in eye tracking regardless of transparency).

morphological-level representations in lexical processing, consistent with a view of lexical representation and processing involving structurally mediated computations over abstract lexical representations.

REFERENCES

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of nonperiodical environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity, 11,* 669–672.

Allen, M., & Badecker, W. (1999). Stem homograph inhibition and stem allomorphy: Representing and processing inflected forms in a mutilevel lexical system. *Journal of Memory and Language, 41,* 105–123.

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25,* 726–740.

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

Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: a group study using MEG. *Neuroreport, 14,* 1183–1187.

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. Booj & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 109–150). Dordrecht: Kluwer Academic.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37,* 94–117.

Badecker, W., & Allen, M. (2002). Morphological parsing and the perception of lexical identity: A masked priming study of stem-homographs. *Journal of Memory and Language, 47,* 125–144.

Balota, D., & Chumbley, J. (1984). Are lexical decisions good measures of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Bauer, L. (1983). *English word-formation.* Cambridge: Cambridge University Press.

Bauer, L. (2001). *Morphological productivity.* Cambridge: Cambridge University Press.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy in lexical access: an MEG study. *Cognitive Brain Research, 24,* 57–65.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48,* 615–634.

Bertram, R., Hyönä, J., & Laine, M. (2000). The role of context in morphological processing: Evidence from Finnish. *Language and Cognitive Processes, 15,* 367–388.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., & Hyönä, J. (2000). Affixal homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language. *Cognition, 74,* B13–B25.

Braeutigam, S., Bailey, A. J., & Swithenby, S. J. (2001). Phase-locked gamma band responses to semantic violation stimuli. *Cognitive Brain Research, 10,* 365–377.

Burani, C., Dovetto, F. M., Spuntarelli, A., & Thornton, A. M. (1999). Morpholexical access and naming: The semantic interpretability of new root-suffix combinations. *Brain and Language, 68,* 333–339.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production: Vol. 2* (pp. 257–294). London: Academic Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*, 297–332.

Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. New York: American Heritage Publishing Co., Inc.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interface environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.

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.

Coolen, R., van Jaarsveld, H. J., & Schreuder, R. (1991). The interpretation of isolated novel nominal compounds. *Memory and Cognition*, *19*, 341–352.

Cornelissen, P., Tarkiainen, A., Helenius, P., & Salmelin, R. (2003). Cortical effects of shifting letter position in letter strings of varying length. *Journal of Cognitive Neuroscience*, *5*, 731–746.

Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes*, *15*, 487–506.

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.

Domínguez, A., Cuetos, F., & Segui, J. (2000). Morphological processing in word recognition: A review with particular reference to Spanish data. *Psicológica*, *21*, 375–401.

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, *53*, 810–842.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*, 301–306.

Embick, D., Hackl, M., Schaeffer, Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, *10*, 345–348.

Eulitz, C., Eulitz, E., Maess, B., Cohen, R., Pantev, C., & Elbert, T. (2000). Magnetic brain activity evoked and induced by visually presented words and nonverbal stimuli. *Psychophysiology*, *37*, 447–455.

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1431–1444.

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, *68*, 33–39.

Fiorentino, R., & Poeppel, D. (2003). *MEG evidence for phrase structure building in noun phrases*. Poster presented at the 33rd Annual Meeting of the Society for Neuroscience, New Orleans, LA, USA, November 8–12.

Fiorentino, R., & Poeppel, D. (2004). Decomposition of compound words: An MEG measure of early access to constituents. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fiorentino, R. (2006). *Masked morphological priming of compound constituents: Specifying the locus of morpho-orthographic segmentation effects using English compound words*. Unpublished raw data.

Forster, K. I. (1988). Basic issues in lexical processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 75–107). Cambridge, MA: The MIT Press.

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, *27*, 203–233.

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language*, *68*, 5–15.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin Company.

Frost, R., Deutsch, A., & Forster, K. I. (2000a). Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 751–765.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. D. (2000b). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory and Cognition*, *28*, 1277–1288.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 829–856.

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language*, *81*, 723–735.

Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*, B27–B41.

Gill, K. M., & McKeever, W. F. (1974). Word length and exposure time effects on the recognition of bilaterally presented words. *Bulletin of the Psychonomic Society*, *4*, 173–175.

Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, *15*, 421–444.

Gonnerman, L., Seidenberg, M. S., & Andersen, E. S. (2006). *Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology*. Unpublished raw data.

Halgren, E., Dhond, R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*, 1101–1116.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Hay, J., & Baayen, H. (2002). Parsing and Productivity. In G. E. Booij & J. van Marie (Eds.), *Yearbook of Morphology 2001* (pp. 203–235). Dordrecht: Kluwer Academic Publishers.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *TRENDS in Cognitive Sciences*, *9*, 342–348.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in left temporal cortex. *Brain*, *121*, 1133–1142.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1999). Semantic cortical activation in dyslexic readers. *Journal of Cognitive Neuroscience*, *11*, 535–550.

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *Journal of Neuroscience*, *22*, 2936–2944.

Henderson, L. (1982). *Orthography and word recognition in reading*. New York: Academic Press.

Hinojosa, J. A., Martín-Loeches, M., Muñoz, F., Casado, P., & Pozo, M. A. (2004). Electrophysiological evidence of automatic early semantic processing. *Brain and Language*, *88*, 39–46.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1612–1627.

Hyönä, J., Vainio, S., & Laine, M. (2002). A morphological effect obtains for isolated words but not for words in sentence context. *European Journal of Cognitive Psychology*, *14*, 417–433.

Inhoff, A. W., Briihl, D., & Schwartz, J. (1996). Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory and Cognition*, *24*, 466–476.

Isel, F., Gunter, T. C., & Friederici, A. D. (2003). Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 277–288.

Jannsen, N., Bi, Y., & Caramazza, A. (2006). *A tale of two frequencies: Determining the speed of lexical access for English and Mandarin Chinese compounds.* Unpublished raw data.

Jarema, G., Busson, C., & Nikolova, R. (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, *68*, 362–369.

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*, *94*, 223–244.

Koyama, S., Kakigi, R., Hoshiyama, M., & Kitamura, Y. (1998). Reading of Japanese Kanji (morphograms) and Kana (syllabograms): A magnetoencephalographic study. *Neuropsychologia*, *36*, 83–98.

Koyama, S., Naka, D., & Kakigi, R. (1999). Evoked magnetic responses during a word completion task. *Electroencephalography and Clinical Neurophysiology*, *49*, 174–178.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Laudanna, A., Cermele, A., & Caramazza, A. (1997). Morpho-lexical representations in naming. *Language and Cognitive Processes*, *12*, 49–66.

Lavidor, M., & Ellis, A. (2002). Word length and orthographic neighborhood size effects in the left and right cerebral hemispheres. *Brain and Language*, *80*, 45–62.

Levelt, W. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*, 13464–13471.

Levi, J. N. (1978). *The syntax and semantics of complex nominals.* New York: Academic Press.

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

Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*, 313–334.

Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory and Cognition*, *5*, 690–695.

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

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*, 487–497.

Martín-Loeches, M., Hinojosa, J. A., Gómez-Jarabo, G., & Rubia, F. A. (1999). The recognition potential: An ERP index of lexical access. *Brain and Language*, *70*, 364–384.

Masson, M., & Bodner, G. E. (2003). A retrospective view of masked priming: Toward a unified account of masked and long-term repetition priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 57–94). New York: Psychology Press.

Masson, M. E. J., & Isaak, M. I. (1999). Masked priming of words and nonwords in a naming task: Further evidence for a nonlexical basis for priming. *Memory and Cognition*, *27*, 399–412.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46.

McQueen, J. M., & Cutler, A. (1998). Morphology in word recognition. In A. Zwicky & A. Spencer (Eds.), *The handbook of morphology* (pp. 406–427). Oxford: Blackwell.

Meunier, F., & Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language*, *68*, 54–60.

Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), *Progress in the psychology of language, Vol. 2* (pp. 147–195). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*, 43–71.

New, B., Brysbaert, M., Segui, J., Ferrand, L., & Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, *51*, 568–585.

Ozawa, K., Fiorentino, R., & Poeppel, D. (2003). *Lexical frequency and the M350 MEG response*. Poster presented at the 33rd Annual Meeting of the Society for Neuroscience, New Orleans, LA, USA, November 8–12.

Plag, I. (1999). *Morphological productivity: structural constraints in English derivation*. New York: Mouton de Gruyter.

Plaut, D.C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*, 445–485.

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–118). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pollatsek, A, & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, *20*, 261–290.

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 820–833.

Pulvermüller, F. (2002). *The neuroscience of language*. Cambridge: Cambridge University Press.

Pylkkänen, L., Feintuch, S., Hopkins, E. M., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: an MEG study. *Cognition*, *91*, B35–B45.

Pylkkänen, L., Llinás, R., & Murphy, G. (2006). Representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, *18*, 1–13.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*, 187–189.

Pylkkänen, L., Stringfellow, A., & Marantz., A. (2002). Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*, 666–678.

Rastle, K., & Davis, M. H. (2003). Reading morphologically complex words: Some thoughts from masked priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 279–305). New York: Psychology Press.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time course study. *Language and Cognitive Processes*, *15*, 507–537.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review*, *11*, 1090–1098.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Roelofs, A., & Baayen, H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin and Review*, *9*, 132–138.

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*, *42A*, 529–567.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition*, *12*, 315–328.

Sekiguchi, T., Koyama, S., & Kakigi, R. (2001). The effect of stimulus repetition on cortical magnetic responses evoked by words and nonwords. *NeuroImage*, *14*, 118–128.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, *25*, 425–437.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, *9*, 2195–2200.

Shoolman, N., & Andrews, S. (2003). Racehorses, reindeer, and sparrows: Using masked priming to investigate morphological influences on compound word identification. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 241–278). New York: Psychology Press.

Simos, P. G., Breier, J. I., Fletcher, J. M., Foorman, B. R., Castillo, E. M., & Papanicolaou, A. C. (2002). Brain mechanisms for reading words and pseudowords: an integrated approach. *Cerebral Cortex*, *12*, 297–305.

Spencer, A. (1991). *Morphological theory*. Cambridge, MA: Blackwell Publishers.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, *14*, 17–26.

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

Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language*, *90*, 88–94.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*, 263–272.

Taft, M. (1991). *Reading and the mental lexicon*. Hove: Lawrence Erlbaum Associates Inc.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, *57A*, 745–765.

Taft, M., & Forster, K. I. (1975). Lexical storage and the retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 17–26.

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., & Kougious, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain and Language*, *90*, 9–16.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*, 151–162.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*, 2119–2132.

Van Jaarsveld, H., & Rattink, G. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, *17*, 447–473.

Van Petten, C. M., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*, 279–293.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.

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

APPENDIX A

Target items (Compounds, single words, and word-nonword foils)

Compound words				Compound word subsets		Foils
AIRPLANE	SNOWFLAKE	KEYWORD	*High frequency*	*Mid frequency*	*Low frequency*	CANTRESK
ARMBAND	SOUTHWEST	LIFERAFT	BASEBALL	BOMBSHELL	BEELINE	CHAIRMIG
BARBELL	TEASPOON	LOGJAM	BATHROOM	BOOKSTORE	CLUBMATE	CROWSKEP
BIRTHDAY	TREETOP	MAILBAG	COWBOY	CREWMAN	FOGHORN	DRABSKEN
BULLFIGHT	BASEBALL	OXTAIL	FORTNIGHT	FANFARE	HUMPBACK	FOOTBAWP
CORNFLAKE	BATHROOM	RAGTIME	GUIDELINE	FOOTPATH	KEYWORD	FRAYGRET
COURTYARD	COWBOY	SOAPBOX	HOUSEHOLD	HAIRCUT	LIFERAFT	FRETSDOP
DOORSTEP	FORTNIGHT	TAPEWORM	LANDMARK	HANDGUN	LOGJAM	HATFOSH
EARLOBE	GUIDELINE	BOMBSHELL	POSTCARD	HEATWAVE	MAILBAG	HILLSIJE
FLAGSHIP	HOUSEHOLD	BOOKSTORE	RAINBOW	LOOPHOLE	OXTAIL	MOUTHPEEM
FLOORMAT	LANDMARK	CREWMAN	SPOTLIGHT	SIDEWALK	RAGTIME	NUTSHEP
GANGLAND	POSTCARD	FANFARE	SUNSHINE	SOYBEAN	SOAPBOX	PANCABE
GIRLFRIEND	RAINBOW	FOOTPATH	WORKSHOP	WOODCHIP	TAPEWORM	POTDASK
HEADACHE	SPOTLIGHT	HAIRCUT				ROPEWAST
NORTHEAST	SUNSHINE	HANDGUN	*Single word subsets*			TRAYBLESH
PAYROLL	WORKSHOP	HEATWAVE	*High frequency*	*Mid frequency*	*Low frequency*	WEARPLATZ
RAILWAY	BEELINE	LOOPHOLE	BASKET	BOUTIQUE	ANDROID	
SEALINK	CLUBMATE	SIDEWALK	CHOCOLATE	KNUCKLE	DERVISH	
SHOWCASE	FOGHORN	SOYBEAN	CREATURE	MIGRAINE	FRISBEE	
SKINCARE	HUMPBACK	WOODCHIP	FRACTION	PALETTE	HARLOT	
			FRAGMENT	PLACARD	HYDRANT	
Single words			FRANCHISE	ROULETTE	KERCHIEF	
CASSETTE	TRESTLE	HYDRANT	GRAMMAR	SARDINE	MASSEUSE	
CHAUFFEUR	TRINKET	KERCHIEF	PLAINTIFF	SEMBLANCE	QUININE	
CHEETAH	TROMBONE	MASSEUSE	PLATFORM	STRETCHER	SPROCKET	
CHIMNEY	TRUNCHEON	QUININE	SANCTION	TEMPLATE	STURGEON	

APPENDIX A (*Continued*)

Compound words				Compound word subsets		Foils
CRESCENT	BASKET	SPROCKET	SEQUENCE	THROTTLE	THIMBLE	
CREVASSE	CHOCOLATE	STURGEON	SUBSTANCE	TURBINE	WOMBAT	
DISCOURSE	CREATURE	THIMBLE				
FOUNTAIN	FRACTION	WOMBAT				
GRIEVANCE	FRAGMENT	BOUTIQUE				
MEMBRANE	FRANCHISE	KNUCKLE				
MERCHANT	GRAMMAR	MIGRAINE				
MISSILE	PLAINTIFF	PALETTE				
PAMPHLET	PLATFORM	PLACARD				
PARLANCE	SANCTION	ROULETTE				
PHEROMONE	SEQUENCE	SARDINE				
PLEASURE	SUBSTANCE	SEMBLANCE				
PRATTLE	ANDROID	STRETCHER				
PRESTIGE	DERVISH	TEMPLATE				
SCOUNDREL	FRISBEE	THROTTLE				
SYNAPSE	HARLOT	TURBINE				

APPENDIX B

Item control statistics

A. Morphemic vs. whole-word properties in the overall comparison of compound vs. single words

Whole compounds and single words were matched on letter length, frequency, and syllabicity (all $t < 1$ by paired two-tailed t-test). Compound words were selected such that morphemic frequency, length, and syllabicity contrasted with the whole compounds and single words. Statistical tests are summarised below.

   a. *Morphemic frequency (constituents higher than whole words).* Analysis of variance (ANOVA) for overall CW, overall SW, C1, and C2 frequency, $F(3, 236) = 104.778$, $MSE = 0.438$, $p < .001$; all planned contrasts: CW vs. C1, CW vs. C2, SW vs. C1, SW vs. C2, $p < .001$.

   b. *Morphemic length (constituents shorter than whole words).* ANOVA for length of overall CW, overall SW, C1, and C2 significant, $F(3, 236) = 525.646$, $MSE = 0.577$, $p < .001$; all planned contrasts: CW vs. C1, CW vs. C2, SW vs. C1, SW vs. C2, significant at $p < .001$.

   c. *Morphemic syllabicity (constituents shorter than whole words).* All constituent morphemes were monosyllabic, and all whole words disyllabic.

B. Morphemic vs. whole-word properties in the comparison of compound vs. single words: Subanalyses at three frequency levels

Compounds and single words were matched on letter length, frequency, and syllabicity (all $t < 1$ by paired two-tailed t-test) within three bins of 12 words each: one for high frequency words, one for middle frequency words, and one for low frequency words. Statistical tests for morphemic vs. whole-word properties are summarised below.

High frequency

   a. *Morphemic frequency (constituents higher than whole words).* ANOVA for the whole-word vs. morpheme frequency manipulation (whole CW, whole SW, C1, and C2) was significant, $F(3, 44) = 23.092$, $MSE = 0.150$, $p < .001$; all planned contrasts: CW vs. C1, CW vs. C2, SW vs. C1, SW vs. C2, significant at $p < .001$.

   b. *Morphemic length (constituents shorter than whole words).* ANOVA for the length mismatch as also significant, $F(3, 44) = 107.498$, $p < .001$; all planned contrasts as above, significant at $p < .001$.

   c. *Morphemic syllabicity (constituents shorter).* All constituents monosyllabic, all whole words disyllabic.

Middle frequency

   a. *Morphemic frequency (constituents higher than whole words).* ANOVA on the frequency mismatch was significant, $F(3, 44) = 51.324$, $MSE = 0.183$, $p < .001$; all planned contrasts significant at $p < .001$.

   b. *Morphemic length (constituents shorter than whole words).* ANOVA showed a significant mismatch, $F(3, 44) = 168.015$, $MSE = 0.373$, $p < .001$; all planned contrasts significant at $p < .001$.

   c. *Morphemic syllabicity (constituents shorter).* All constituents monosyllabic, all whole words disyllabic.

APPENDIX B (*Continued*)

Low frequency

  a. *Morphemic frequency (constituents higher than whole words).* ANOVA on the frequency mismatch was significant, $F(3, 44) = 50.397$, $MSE = 0.354$, $p < .001$; all planned contrasts significant at $p < .001$.

  b. *Morphemic length (constituents shorter than whole words).* ANOVA showed a significant mismatch, $F(3, 44) = 102.898$, $MSE = 0.481$, $p < .001$; all planned contrasts significant at $p < .001$.

  c. *Morphemic syllabicity (constituents shorter)*. All constituents monosyllabic, all whole words disyllabic.

Additional tests of letter-length across frequency bins

Whole-word length across frequency bins. The compounds and single words were matched identically for length within each bin. There was a small length difference across bins, $F(2, 33) = 4.092$, $MSE = 0.631$, $p < .027$; only the high- and low-frequency items differed significantly in a planned contrast, $t(33) = 2.826$, $p < .009$.

Morphemic length across frequency bins. Within the high frequency bin, first and second compound constituent letter lengths were 4.00 and 4.08 ($t < 1$), within medium frequency bin 3.42 and 3.75 ($t < 1$), and within low frequency 3.83 and 3.92 ($t < 1$). First constituent lengths differed by a fraction of a letter across frequency bins, $F(2, 33) = 3.906$, $MSE = 0.348$, $p < .031$; only medium and low frequency first-constituents differed significantly in a planned contrast, $t(33) = 2.766$, $p < .01$. Second constituent lengths did not differ significantly across frequency bins.

APPENDIX C
Results and statistical tests, psychophysical replication study

TABLE 5A
Response time and accuracy ($N = 12$) for replication study

	Response time (Acc %)
Overall comparison	
Compound words	605 (92%)
Single words	678 (82%)
Word-nonword foils	722 (86%)
Subanalyses at three frequency levels	
High frequency	
Compound words	557 (98%)
Single words	623 (99%)
Mid frequency	
Compound words	609 (96%)
Single words	690 (88%)
Low frequency	
Compound words	667 (81%)
Single words	766 (67%)

TABLE 5B
Statistical tests for Table 5A

	Response time		Accuracy	
	ANOVA by-participants	*ANOVA by-items*	*ANOVA by-participants*	*ANOVA by-items*
Overall comparison				
CW vs. SW. vs. Foil	$F_1(2, 22) = 24.856, p < .001$	[a]	$F_1(2, 22) = 10.901, p < .002$	[a]
CW vs. SW	$F_1(1, 11) = 20.182, p < .002$	$F_1(1, 59) = 23.324, p < .001$	$F_1(1, 11) = 18.140, p < .002$	$F_2(1, 59) = 11.934, p < .001$
CW vs. Foil	$F_1(1, 11) = 35.487, p < .001$	[a]	$F_1(1, 11) = 7.096, p < .023*$	[a]
Subanalyses at three frequency levels				
Structure (CW vs. SW)	$F_1(1, 11) = 17.786, p < .002$	$F_2(1, 11) = 15.596, p < .003$	$F_1(1, 11) = 11.249, p < .007$	$F_2(1, 11) = 9.888, p < .010$
Frequency (high, mid, low)	$F_1(2, 22) = 34.125, p < .001$	$F_2(2, 22) = 11.479, p < .001$	$F_1(2, 22) = 47.680, p < .001$	$F_2(2, 22) = 20.318, p < .001$
Interaction	$F_1(2, 22) = 0.610, p < .553^{\dagger}$	$F_2(2, 22) = 0.273, p < .764^{\dagger}$	$F_1(2, 22) = 6.156, p < .009$	$F_2(2, 22) = 5.236, p < .015$
High vs. mid	$F_1(1, 11) = 16.504, p < .002$	$F_2(1, 11) = 14.416, p < .004$	$F_1(1, 11) = 14.865, p < .004$	$F_2(1, 11) = 16.036, p < .003$
Interaction	$F_1(1, 11) = 0.333, p < .577^{\dagger}$	$F_2(1, 11) = 0.583, p < .462^{\dagger}$	$F_1(1, 11) = 10.170, p < .010$	$F_2(1, 11) = 7.932, p < .018$
Mid vs. low	$F_1(1, 11) = 30.323, p < .001$	$F_2(1, 11) = 14.416, p < .004$	$F_1(1, 11) = 50.005, p < .001$	$F_2(1, 11) = 12.558, p < .006$
Interaction	$F_1(1, 11) = 0.265, p < .618^{\dagger}$	$F_2(1, 11) = 0.483, p < .502^{\dagger}$	$F_1(1, 11) = 2.647, p < .133^{\dagger}$	$F_2(1, 11) = 0.533, p < .474^{\dagger}$
CWH vs. SWH	$F_1(1, 11) = 12.581, p < .006$	$F_2(1, 11) = 10.714, p < .008$	$F_1(1, 11) = 2.200, p < .167^{\dagger}$	$F_2(1, 11) = 0.550, p < .475$
CWM vs. SWM	$F_1(1, 11) = 8.037, p < .017$	$F_2(1, 11) = 8.483, p < .015$	$F_1(1, 11) = 8.250, p < .016$	$F_2(1, 11) = 6.838, p < .025$
CWL vs. SWL	$F_1(1, 11) = 11.063, p < .008$	$F_2(1, 11) = 6.352, p < .029$	$F_1(1, 11) = 8.105, p < .017$	$F_2(1, 11) = 8.737, p < .014$

[†] n.s.

[a] Foils not entered into by-items due to the large difference in number of samples.

[*] n.s. in main experiment.

TABLE 5C
Reanalysis of 5B with six items removed

The following analyses were conducted after six items were removed to reduce the possibility that some single words were also treated as complex (grievance, stretcher, creature, substance, merchant, pleasure). The pattern of results is virtually the same as in Table 5B. Differences (three cases in which a statistical result became marginal rather than significant) are marked with the †† mark, as indicated below the table.

	Response time		Accuracy	
	ANOVA by-participants	*ANOVA by-items*	*ANOVA by-participants*	*ANOVA by-items*
Overall comparison				
CW vs. SW. vs. Foil	$F_1(2, 22) = 26.845, p < .001$ [a]		$F_1(2, 22) = 13.932, p < .001$ [a]	
CW vs. SW	$F_1(1, 11) = 26.138, p < .001$	$F_2(1, 59) = 29.732, p < .001$	$F_1(1, 11) = 20.822, p < .002$	$F_2(1, 59) = 13.549, p < .002$
CW vs. Foil	$F_1(1, 11) = 35.487, p < .001$ [a]		$F_1(1, 11) = 7.096, p < .023^*$ [a]	
Subanalyses at three frequency levels				
Structure (Cw vs. SW)	$F_1(1, 11) = 20.228, p < .002$	$F_2(1, 11) = 16.776, p < .003$	$F_1(1, 11) = 11.807, p < .007$	$F_2(1, 11) = 7.832, p < .018$
Frequency (High, mid, low)	$F_1(2, 22) = 36.575, p < .001$	$F_2(2, 22) = 11.053, p < .001$	$F_1(2, 22) = 46.448, p < .001$	$F_2(2, 22) = 20.520, p < .001$
Interaction	$F_1(2, 22) = 0.427, p < .659^\dagger$	$F_2(2, 22) = 0.134, p < .875^\dagger$	$F_1(2, 22) = 5.949, p < .009$	$F_2(2, 22) = 4.313, p < .027$
High vs. Mid	$F_1(1, 11) = 19.603, p < .002$	$F_2(1, 11) = 15.314, p < .003$	$F_1(1, 11) = 15.223, p < .003$	$F_2(1, 11) = 12.077, p < .006$
Interaction	$F_1(1, 11) = 0.092, p < .768^\dagger$	$F_2(1, 11) = 0.274, p < .612^\dagger$	$F_1(1, 11) = 10.727, p < .008$	$F_2(1, 11) = 4.592, p < .056^{\dagger\dagger}$
Mid vs. Low	$F_1(1, 11) = 29.275, p < .001$	$F_2(1, 11) = 4.719, p < .054^{\dagger\dagger}$	$F_1(1, 11) = 47.651, p < .001$	$F_2(1, 11) = 14.862, p < .004$
Interaction	$F_1(1, 11) = 0.299, p < .597^\dagger$	$F_2(1, 11) = 0.008, p < .932^\dagger$	$F_1(1, 11) = 2.107, p < .176^\dagger$	$F_2(1, 11) = 1.505, p < .247^\dagger$
CWH vs. SWH	$F_1(1, 11) = 16.263, p < .003$	$F_2(1, 11) = 16.263, p < .003$	$F_1(1, 11) = 1.678, p < .223^\dagger$	$F_2(1, 11) = 0.305, p < .593$
CWM vs. SWM	$F_1(1, 11) = 9.621, p < .011$	$F_2(1, 11) = 9.621, p < .011$	$F_1(1, 11) = 8.932, p < .013$	$F_2(1, 11) = 3.713, p < .081^{\dagger\dagger}$
CWL vs. SWL	$F_1(1, 11) = 11.063, p < .008$	$F_2(1, 11) = 11.063, p < .008$	$F_1(1, 11) = 8.105, p < .017$	$F_2(1, 11) = 7.139, p < .023$

† n.s.

†† Marginal here, significant in main and replication experiments.

[a] Foils not entered into by-items due to the large difference in number of samples.

* n.s. in main experiment.