WHOLE-WORD FREQUENCY EFFECTS IN ENGLISH MASKED PRIMING: VERY LITTLE *CORN* IN *CORNER* AND *CORNET*

Kaidi Lõo[1] and Juhani Järvikivi[2]

1 – Department of Psychology, University of Alberta
2 – Department of Linguistics, University of Alberta
kloo@ualberta.ca, jarvikivi@ualberta.ca

ABSTRACT

The question whether complex words, including pseudo-complex words (e.g., *corn+er*), are obligatorily segmented into existing morphemes (e.g., [24]) has been the topic of a large body of past morphological processing research. A recent line of studies finds consistent effects of the whole-word already early on in the processing (e.g., [11,20]), challenging the obligatory decomposition view. In our current masked priming study with native speakers of English, participants showed facilitation for true morphological relations only as neither *corner* or *cornet* words produced significant priming. Additionally, frequency of the target and the prime affected processing of real and pseudo morphology differently.

Keywords: whole-word frequency, morphological processing, masked priming

1. INTRODUCTION

Proponents of obligatory decomposition (see e.g., [3,13-15,19,22-24]) argue that all complex words must first undergo a form-based morphological decomposition. In this approach, effects reflecting the whole complex word, including semantic transparency and whole-word frequency effects, can only be explained as late effects, possibly reflecting integration processes occurring at or after the recombination of morphemes [10,27].

The strong view of obligatory decomposition has found support from a line of masked priming studies. These studies show comparable facilitation for words with real decomposable and only apparent structure like *harpist* and *corner*, but not for words with just embedded stems like *harpoon* or *cornet* [24], suggesting automatic parsing based on form alone.

These results are in contrast with research suggesting that various semantic effects pertaining to the whole complex word form affect recognition at various points during the course of processing, even early on [5-9,11-12,25].

Rueckl and Aicher [25] found no effects in long lag priming for pseudo-complex words, unlike for transparent complex words, suggesting that at the very least pseudo structure is handled differently than real morphological relationships are. Järvikivi et al. [12] reported that priming for pseudo-derived words was smaller than for derived words. Järvikivi and Pyykkönen [11] found that pseudo-complex words resulted in signicantly less priming than inflected words. They also reported that the morphological family size inhibited processing when the family size of the pseudo-complex prime was larger than the family size of the target, suggesting an early influence of semantics.

Further, Feldman and colleagues [8] showed that if one controls for the stem and affix properties between the different conditions, semantically transparent pairs show stronger priming than opaque pairs in English, and the same was later shown for Serbian [6]. In another study, Feldman et al. [9] varied the stimulus onset asynchrony (SOA) between the target and the prime for semantically similar and dissimilar pairs, and showed that semantic priming increased linearly with increasing SOA. Importantly, reaction times were already signicantly faster for semantically related pairs at SOAs of 34 ms and 48 ms.

Schmidtke et al. [26] conducted a distributional survival analysis of lexical decision latencies for English and Dutch derived words and found that whole-word frequency emerged the earliest (419 ms), followed by stem (i.e., morphemic constituent) frequency (437 ms). Furthermore, a recent priming study by [20] found comparable priming effects for pseudo-derived words (e.g., *corner*) and orthographic controls (e.g., *brothel*). They further showed that more experienced readers are affected by priming to a lesser degree compared to less experienced readers.

Marelli et al. [18] found no effects of morphological segmentation for opaque words in Italian, when using a semantic task instead of priming task, but an item set equivalent to those typically used in priming studies.

Taken together with recent studies showing paradigm size and whole-word frequency effects for Estonian inflected words (see [16-17], for word recognition and production, respectively), these studies suggest that information about whole word form, and not only about morphemic constituents, is relevant in online language processing, and possibly very early on.

Furthermore, most masked priming studies have not only been done between-item but have compared pseudo-derived words like *corner* to true derived words like *dancer*. This might also explain why the effects of equal magnitude for the pseudo-complex words and real complex words have been interpreted as involving similar processing. In the current study, we asked the extent to which this is the case by comparing the effects of prime and target distributional properties, across the prime types within-item for native speakers of English.

2. EXPERIMENT

2.1. Participants

77 native speakers of English (59 female, mean age 21 years, range 18-44) with normal or corrected-to-normal vision from the University of Alberta participated in the experiment for partial course credit.

2.2. Materials

Forty English words were selected as target stimuli from the Corpus of Contemporary American English (COCA). Each word target (e.g., *corn*) was primed within-item in four conditions: transparent (e.g., *corns*), opaque (e.g., *corner*), stem-embedded/form (e.g., *cornet*), and unrelated control (e.g., *eyes*). We used both derived and inflected forms of the stem in the transparent condition. Additionally, 60 nonwords and 20 further real words were added to the item set as fillers. Nonword targets (e.g., *shoop*). were created by changing one or two letters from an existing English word. Primes for these words were always real English words, consisting of the same four types with the same proportions as for the real word targets.

2.3. Design and procedure

The prime-target pairs were counterbalanced across four lists. Each list contained 120 pairs, including 40 experimental prime-target trials and 80 filler trials (with 20 real word and 60 non-word targets). The experiment was carried out using the E-Prime experimental software (Psychology Software Tools Inc.) and SRBOX response box. All stimuli were presented in black 32-point font Courier New letters on light gray background at the centre of the computer screen.

Each trial began with a fixation cross (+) set to appear in the centre of the screen for 1000 ms, immediately followed by a forward mask (#####) for 500 ms. After that, the prime word appeared in lower case letters in the same location for 50 ms. Finally, the target word appeared in the same location in upper case letters, and remained on the screen until the participant decided whether the word on the screen is an existing English word or not by pressing a relevant button on the response box. The participant was instructed to respond as fast and as accurately as possible. Experimental trials were preceded by 10 practice trials. The experiment lasted approximately 20 minutes.

3. ANALYSIS AND RESULTS

First, nonword trials and fillers were removed from the dataset. Trials with response times less than 50 ms and more than 1500 ms (2.8% of the data) as well as trials with incorrect responses (12.1% of the data) were also removed from the response time analysis.

Frequencies for the primes and targets were determined using COCA [4], and morphological family size for the primes and targets using CELEX database [1]. The stem frequency and whole-word frequency was the same for the target words, as they were monomorphemic (e.g., *corn*). For primes, whole-word frequency (i.e., the token frequency of e.g., *corns*, *corner* or *cornet*) was used. Both frequency and morphology family size were log-transformed prior the analysis to reduce the skewness of their distributions.

The results of the response time analysis were analysed using Generalized Additive Mixed Models [28] with R-package *mgcv* [2] in two steps.

In the first step, only the effect of condition was investigated. The first model included condition as a predictor and by-subject factor smooths for trial and by-target random intercepts as random effects. Factor smooths allowed for the shape of the average distribution of reaction times across the experiment to vary by-subject. Participants showed significant facilitation for the transparent words (t=-23.24, p=0.004), whereas the other two conditions did not differ from the unrelated condition ($p>0.05$).

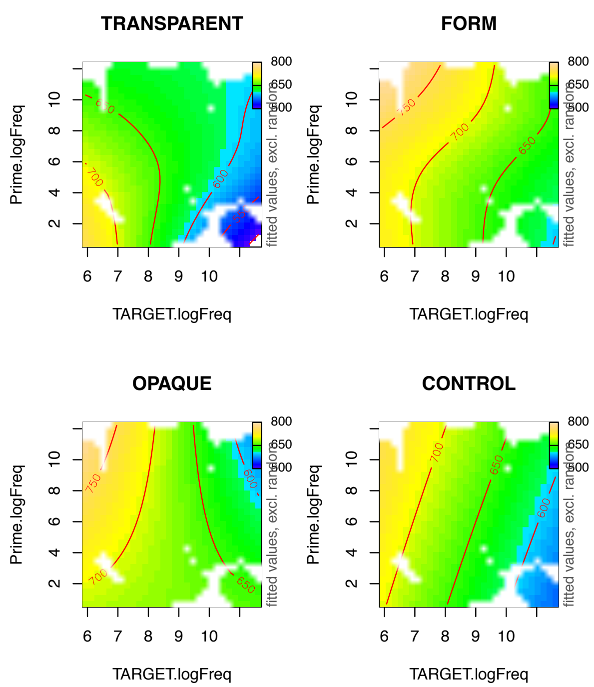
Table 1: Mean (M) and Standard deviation (SD) of reaction times by condition in milliseconds.

Condition	Transparent	Form	Opaque	Contol
M	668.76	711.64	715.49	706.70
SD	198.58	217.75	226.27	225.53

In the second step, the frequency and morphological family size of the target and prime, and their interactions between them were also added to the model. Random structure of the model remained the same. In this model, transparent condition was still processed signicantly faster than unrelated condition (t=-2.33, p=0.02), however, now form condition became significantly slower compared to unrelated condition (t=2.77, p=0.006), whereas morphologically opaque condition remained indifferent from unrelated condition (t=1.5, p=0.13).

Importantly, adding frequency of the target and the prime to the model showed a signicant interaction for condition. However, the effect was different for the transparent condition compared to the form and opaque conditions. In the transparent condition, responses to low prime and low target frequency were the slowest. In the other two conditions, responses with higher prime than target frequency were the slowest. This is shown by Figure 1, where colour coding is used to represent model predictions, with darker shades of yellow indicating slower reaction times, and darker shades of blue representing faster reaction times.

Figure 1: Tensor product smooth for the interaction of prime and target word frequency by condition.



The family size ratio between prime and target was not significant, possibly because of the relatively small family sizes in English (cf. [21], for Finnish).

In summary, our analysis showed a significant facilitation for the transparent words, but not for the pseudo-complex and stem-embedded words. The investigation of prime and target frequency interactions indicated between condition differences: in transparent condition responses with low prime and target frequency were the slowest, in the pseudo-complex and stem-embedded conditions responses with higher prime than target frequency were the slowest.

4. DISCUSSION AND CONCLUSION

The current study investigated early morphological processing of English using the masked priming paradigm. Our results are in line with other recent evidence finding different effects of real and pseudo morphology, with respect to distributional properties in masked priming [8,11,20]. The present study expanded the existing research in three ways.

First, we extended the research indicating that native speakers of English show facilitation for true morphological relations only. Not only were the responses for *corner* and *cornet* words the same in our study, but neither produced signicant priming. These results are inline with findings by Rueckl and Aicher [25] as well as Järvikivi and Pykkönen [11], who also found less or no priming for pseudo-complex words, compared to words with real morphological structure.

Second, the prime-target frequency ratio inhibited processing when the primes were higher frequency than the targets, but only for *corner* and *cornet* words, and to the same degree. This is in line with Milin et al. [20] who also found comparable priming for these two types of words in priming using learning-based measures rather than lexical-distributional measures like in the current study. This is further evidence suggesting that the underlying processing for both these word types differs from words with real morphological structure.

Third, we used a complete within-item design in the masked priming paradigm, including primes across conditions with the same (pseudo)stem (but see also [20]), making the comparison between true and pseudo-morphology more precise.

In summary, the present study extended previous research by applying within-item design and accounting for various properties of the target and the prime in order to investigate morphological processing in masked priming in more detail than many previous masked priming studies.

Future research will need to address the extent to which past evidence for automatic blind segmentation of morphologically complex words is an artifact of the restricted set of materials usually used in this task and to what extent it is an actual morphological processing effect. More studies using more diverse language populations as well as applying within-item design in masked priming are needed to get to the bottom of the exact nature of morphological processing.

REFERENCES

1. Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. *Linguistic data consortium*. University of Pennsylvania, Philadelphia.
2. Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94, 206-234.
3. Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, 23(2), 533-539.
4. Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available online at https://corpus.byu.edu/coca/.
5. 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(6), 1431-1444.
6. Feldman, L. B., Kostić, A., Gvozdenović, V., O'Connor, P. A., & del Prado Martín, F. M. (2012). Semantic similarity influences early morphological priming in Serbian: A challenge to form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 19(4), 668-676.
7. Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, 9, 111.
8. Feldman, L. B., O'Connor, P. A., & del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4), 684-691.
9. Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, 68(1-2), 33-39.
10. Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain and Language*, 143, 81-96.
11. Järvikivi, J., & Pyykkönen, P. (2011). Sub-and supralexical information in early phases of lexical access. *Frontiers in Psychology*, 2, 282.
12. Järvikivi, J., Pyykkönen, P., & Niemi, J. (2009). Exploiting degrees of inflectional ambiguity: Stem form and the time course of morphological processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 221.
13. Kazanina, N. (2011). Decomposition of prexed words in Russian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1371.
14. Lazaro, M., Illera, V., & Sainz, J. (2016). The sux priming effect: Further evidence for an early morphoorthographic segmentation process independent of its semantic content. *The Quarterly Journal of Experimental Psychology*, 69(1) 197-208.
15. Longtin, C., Segui, J., & Halle, P. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*. 18, 313-334.
16. Lõo, K., Järvikivi, J., & Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, 175, 20-25.
17. Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., & Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology,* 28(1), 71-97.
18. Marelli, M., Amenta, S., Morone, E. A., & Crepaldi, D. (2013). Meaning is in the beholder's eye: Morpho-semantic effects in masked priming. *Psychonomic Bulletin & Review*, 20(3), 534-541.
19. Marslen-Wilson, W. D., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, 23(3), 394-421.
20. Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PloS One*, 12(2), e0171935.
21. Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: the case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1271.
22. Rastle, K. & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23 (7-8), 942-971.
23. 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(4-5), 507-537.
24. Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review,* 11, 1090-1098.
25. Rueckl, J. G. & Aicher, K. A. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. Language and Cognitive Processes, 23(7-8), 972-1001.
26. Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1793-1820.
27. Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, 57(4), 745-765.
28. Wood, S. N. (2006). Generalized additive models: an introduction with R. Chapman and Hall/CRC.