LANGUAGE AND COGNITIVE PROCESSES
2008, 23 (7/8), 1191–1223

Psychology Press
Taylor & Francis Group

A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds

Niels Janssen
CNRS & Université de Provence, Marseille, France, and Harvard University, Cambridge, MA, USA

Yanchao Bi
Beijing Normal University, Beijing, P. R. China, and Harvard University, Cambridge, MA, USA

Alfonso Caramazza
Harvard University, Cambridge, MA, USA, and Center for Mind/Brain Sciences, University of Trento, Italy

Two picture naming experiments show that compound word production in Mandarin Chinese and in English is determined by the compound's whole-word frequency, and not by its constituent morpheme frequency. Four control experiments rule out that these results are caused by recognition or articulatory processes. These results are consistent with models of lexical access that assume compounds are stored in their full-form and that frequency affects the retrieval

---

Correspondence should be addressed to Niels Janssen, Laboratoire de Psychologie Cognitive, CNRS & Université de Provence, 3, place Victor Hugo – Case 66, 13331 Marseille Cedex 3, France. E-mail: janssen@up.univ-mrs.fr

Correspondence can also be addressed to Yanchao Bi, Institute of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, P. R. China. E-mail: yanchao@gmail.com

The research reported here was supported by NIH Grant DC04542 to Alfonso Caramazza. Experiment 1a and 1b were presented at the International Conference of Cognitive Sciences in Beijing, China, August, 2001 by Yanchao Bi. Experiment 2a and 2b were presented at the Trends in Cognitive Psychology Symposium, in Naples, Italy, March, 2004 by Niels Janssen. Niels Janssen was supported by the Sackler Scholar Programme in Psychobiology and a post-doctoral fellowship from the Fyssen Foundation. We thank Jorge Almeida, Raymond Bertram, and one anonymous reviewer for helpful comments and discussion.

http://www.psypress.com/lcp DOI: 10.1080/01690960802250900

of whole words. The present results corroborate the results from previous studies that have investigated compound word production in Mandarin Chinese, but also differ from those previously reported on compound word production in Dutch. The possibility that this inconsistency arises due to cross-linguistic, or task differences is discussed.

Keywords: Compounds; Frequency; Language production; Morphology; Picture naming.

Many languages contain words that are composed out of two or more other words: Compounds. In English, for example, the compound word 'windmill' consists of two existing words 'wind' and 'mill'. In the English language such words are fairly common as can be estimated by consulting the CELEX lexical database, which contains about 4800 (token) occurrences of nominal compounds. For these lexicalised nominal compounds, the question arises how they are stored in the mental lexicon. There are at least two possible hypotheses. One is the *decomposition hypothesis*, which assumes that there is at least one level of lexical representation at which compounds are represented in terms of their constituent lexical morphemes. The alternative view, the *full-form representation hypothesis* (or full-listing hypothesis), assumes that only whole-word forms are represented in the lexicon and that a word's derivational history plays no role in lexical access (see, Butterworth, 1983; Laudanna, Badecker, & Caramazza, 1992).

Levelt, Roelofs, and Meyer (1999) have proposed a model of lexical access in language production that explicitly incorporates the decomposition hypothesis for compounds. The model assumes that there are two levels of lexical representation intervening between the semantic and phonological contents of words: the lemma and lexeme levels (see also Dell, 1986; Garrett, 1980). The lemma level (also called the syntactic stratum) represents the grammatical properties of words (e.g., grammatical gender), whereas the lexeme level represents the phonological form of lexical and grammatical morphemes of the language. In this model, a word such as 'windmill' is represented as a whole word ([windmill]) at the lemma level, and it is represented in terms of its constituent morphemes – /wind/ and /mi:l/ – at the lexeme level. A schematic representation of this architecture is shown in Figure 1a.

The decomposition hypothesis is incompatible with models of lexical access that assume that there is only a single lexical layer between the semantic and phonological contents of words (e.g., Caramazza, 1997). Such models assume that lexicalised compounds are represented as whole words and that their constituent morphemes do not play a role in lexical access (Figure 1b). Thus, evidence showing that compounds are represented in terms of their constituent morphemes would undermine models of language

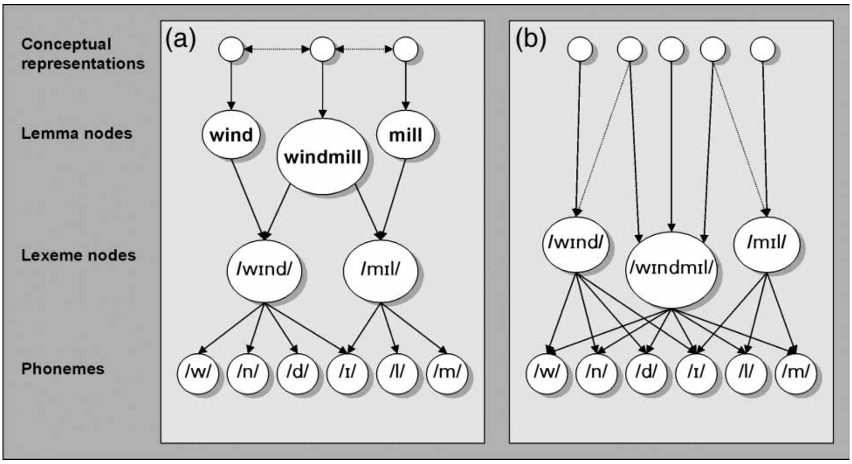

Figure 1. Schematic depiction of compound representation in two-stage (a) and single-stage (b) models of lexical access.

production that assume that there is only a single lexical layer between the semantic and the phonological contents of words.[1]

In order to distinguish between the decomposition and full-form representation hypotheses, one could consider how a manipulation of word frequency affects the processing of compounds. Manipulation of word frequency has a long history in language production research investigating picture naming (e.g., Oldfield & Wingfield, 1965). It is generally believed that frequency affects the retrieval of the picture's name and not the recognition of the picture (e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Griffin & Bock, 1998; Jescheniak & Levelt, 1994; Wingfield, 1967, 1968; but see Kroll & Potter, 1984). The effects of word frequency on speech production has provided information on various aspects of the lexical access process (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Dell, 1990; Griffin & Bock, 1998; Jescheniak & Levelt, 1994). However, the interpretation of such effects is not straightforward but depends on the combination of assumptions one makes about the architecture and processing dynamics of the lexical system, the locus of the frequency effect in the system, and the way in which frequency is assumed to affect processing.

---

[1] Note that the reverse is not true. That is, evidence against the decomposition hypothesis of compounds would not constitute evidence against the lemma/lexeme distinction. This is because the full-form representation hypothesis can be implemented in a lexical access system that distinguishes between lemma and lexeme lexical layers by assuming that compounds are represented as whole words at both levels of representation. That is, lexical models that assume a distinction between lemma and lexeme levels are compatible with both the decomposition and the full-form hypotheses of compounds.

Consider the decomposition model proposed by Levelt et al. (1999). In this model it is assumed that compounds are stored in terms of their constituent morphemes at the lexeme level, and that the frequency variable affects the retrieval of lexeme representations (Jescheniak & Levelt, 1994). On the basis of these assumptions one would expect that the frequency of occurrence of the morphemes constituting the compound would be the principal determinants of naming latency. For example, naming latency in the production of the compound 'windmill' should be determined by the frequency of the morpheme constituents 'wind' and 'mill' – a *morpheme frequency effect* – and not the frequency of the compound as a whole.[2] Alternatively, according to the full-form model proposed by Caramazza (1997), compounds are represented in their full-form at the lexical node level. If it is further assumed that frequency had its effect at this level, one would expect that it is the frequency of occurrence of the compound word that affects naming latencies. For example, for the production of the compound 'windmill' it would be the frequency of occurrence of the word 'windmill' that would determine lexical retrieval times – a *whole-word frequency effect*.

The question that we addressed in this paper is whether the production of compounds is sensitive to morpheme or whole-word frequency. Previous studies that have investigated this issue have not found consistent results. First, a morpheme frequency effect was found in the study of Roelofs (1996a). In this study, participants first learned small sets of associated word pairs (e.g., douche/schuimbad [shower/foam bath]); in the experimental trials the first word of the learned pairs was presented on the computer screen and the participants were asked to produce the second word of the pair (see also Meyer, 1990, 1991). All the target words were Dutch nominal bisyllabic compounds, and the frequency of the first constituent of the compound was either high, or low frequency. In addition, compound production took place in two contexts: In a homogeneous context where compounds produced on consecutive trials shared the initial constituent, or in a heterogeneous context where compounds produced on consecutive trials did not share the initial constituent. The results showed that there was a main effect of context (89 ms facilitation), presumably reflecting the advantage that accrues from being able to prepare the overlapping constituents in the homogeneous sets as opposed to the heterogeneous sets. More important for the present purposes, the context effect was larger when the response words shared

---

[2] Note that with additional assumptions about incremental language production different predictions could be made about the size of frequency effects for the first and second morpheme. However, independent of these assumptions, the model predicts that RTs should decrease with increasing average morpheme frequency.

low-frequency (99 ms) versus high-frequency constituents (79 ms). The larger facilitation observed for low- than high-frequency constituents was explained by assuming that low-frequency constituents take longer to encode and therefore can benefit more from being in the homogeneous condition. On the basis of these results, Roelofs argued that compound word production is sensitive to morpheme frequency.

A similar pattern of results was found in a study by Bien, Levelt, and Baayen (2005). The design in the study of Bien et al was similar to that used by Roelofs (1996a). Like Roelofs' study, participants were native speakers of Dutch, compounds were binominal multi-syllabic compounds, and target production took place in the context of a task in which responses were produced on the basis of associated cues. However, unlike the study of Roelofs, the context variable was not manipulated. In four experiments the frequency of the head of the compound (most Dutch compounds are right-headed), the frequency of the modifier, the frequency of both the head and the modifier, and the whole-word frequency was manipulated. The results revealed various morpheme frequency effects: of the head (14 ms), the modifier (25 ms), the head and the modifier (27 ms), but not of the whole-word frequency.

In contrast to these two studies, a study by Chen and Chen (2006) showed that compound word production in Mandarin Chinese was not sensitive to morpheme frequency. The paradigm in their study was identical to the one used by Roelofs (1996a). In their Experiment 4, the morpheme frequency of the compounds' first constituent was manipulated, while the whole-word frequency of the compounds was controlled. Also, a manipulation of the context variable was included. Like Roelofs' study, the results revealed an overall effect of context (42 ms), but, unlike the results of Roelofs, the effect of context was not modulated by the frequency of the constituents. The context effect was as large when compounds shared low frequency constituents (43 ms) as when compounds shared high frequency constituents (42 ms). Thus, the absence of a morpheme frequency effect in this study stands in contrast to the results of Roelofs (1996a) and Bien et al. (2005).

Deviating from the pattern of results obtained by Roelofs (1996a) and Bien et al. (2005) as well as Chen and Chen (2006), are the results from studies on compound recognition in the field of visual word recognition. These studies reliably report that compound word recognition is a function of both morpheme and whole-word frequency (e.g., De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002; Juhasz, Starr, Inhoff, & Placke, 2003; Libben, Gibson, Yoon, & Sandra, 2003; Pollatsek, Hyönä, & Bertram, 2000; Taft & Zhu, 1997; van Jaarsveld & Rattink, 1988; Zwitserlood, 1994). For example, in a study by Pollatsek et al. (2000), eye fixations of Finnish participants were sensitive to the compounds' morpheme frequency and also to its whole-word frequency. These results have been interpreted in terms of a model that assumes parallel dual-routes of decomposed and whole-word access during visual word recognition (Schreuder & Baayen, 1995).

Finally, the nature of compound word representation has not only been investigated by an assessment of the sensitivity of compound word production to whole-word or morpheme frequency, but also by investigating its sensitivity to priming effects. In a series of studies by Zwitserlood and colleagues (Dohmes, Bölte, & Zwitserlood, 2004; Zwitserlood, Bölte, & Dohmes, 2000, 2002), two variants of a picture-word priming paradigm were employed. One was an immediate variant in which the prime word and target picture appeared simultaneously on a given trial, and the other a delayed variant in which the prime word preceded the presentation of the target picture by at least 7 trials. In both the immediate and delayed naming tasks, target pictures had monomorphemic names (e.g., Berg [mountain]), and prime words were morphologically related compound words (e.g., Berghütte [mountain cabin]), form-related (e.g., Berber [berber]), semantically related (e.g., Hügel [hill]), or unrelated (e.g., Note [mark]). The study was conducted in German. The results revealed that morphological primes facilitated picture naming in both the immediate (142 ms) and delayed tasks (42 ms), while the phonological (17 ms) and semantic primes (–72 ms) affected naming latencies only in the immediate task (see Gumnior, Bölte, & Zwitserlood, 2006 for similar findings in a word-translation paradigm). On the basis of these results, Zwitserlood and colleagues argued for a decomposed representation of compounds and rejected the full-form hypothesis.

However, these results do not rule out that the morphological priming effect observed in their experiments arises at a semantic level. Their argument that this effect reflects priming of a morphological and not a semantic representation rests primarily on the observation of a morphological priming effect in the immediate and delayed tasks, and a semantic priming effect only in the immediate task (Zwitserlood et al., 2000). If the morphological priming effect were semantic in nature, one would expect a similar pattern of semantic and morphological priming effects in both tasks. Thus, the absence of the semantic priming effect in the delayed task presumably rules out a semantic interpretation of the results. However, the observation that there was no semantic priming effect in the delayed task is difficult to reconcile with the results from other studies that have consistently reported semantic priming effects in delayed tasks (e.g., Becker, Moscovitch, Behrmann, & Joordens, 1997; Damian & Als, 2005; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Wheeldon & Monsell, 1992). These results complicate the interpretation of Zwitserlood et al.'s findings, and undermine a clear interpretation of their data in terms of morphological priming.

Thus, the nature of compound word representation, and the sensitivity of compound word production to whole-word or morpheme frequency remain unresolved. Some studies have reported morpheme and not whole-word frequency effects (Bien et al., 2005; Roelofs, 1996a), while in another no morpheme frequency effect was found (Chen & Chen, 2006). The present experiments sought to provide additional evidence on this issue. One possible explanation for the inconsistency in the results between those of Chen and Chen and those of Bien et al. and Roelofs is in terms of cross-linguistic differences. In the former study participants were native speakers of Mandarin Chinese, while in the latter two studies they were native speakers of Dutch. As argued by Chen and Chen, it could be that in languages with a relatively simple morphology, such as Mandarin Chinese, morphological representations play less of a role compared to languages with a relatively complex morphology such as Dutch. If such were the case, one might expect that naming latencies to compounds would reveal no morpheme frequency effect in Mandarin Chinese, and a morpheme frequency effect in Dutch.

In the experiments reported here, participants were native speakers of Mandarin Chinese in Experiment 1, while they were native speakers of English in Experiment 2. In both experiments the whole-word and morpheme frequency properties of the compounds was manipulated. If cross-linguistic differences affect compound word representation, and if morphological complexity in English is comparable to Dutch, one would expect naming latencies in Experiment 1 to be sensitive to whole-word but not morpheme frequency, and naming latencies in Experiment 2 to be sensitive to morpheme but not whole-word frequency.[3]

Finally, compound word production in both Experiment 1 and 2 took place in the context of a picture naming task. This task is different from the response association tasks that have been employed by previous studies on compound word production. Like the response association task, the picture naming task has a long tradition in language production research (e.g., Potter & Faulconer, 1975). In the present study, the question of interest was

---

[3] It has been demonstrated that frequency and age of acquisition (AoA) are highly correlated (see Carroll & White, 1973b; Ellis & Morrison, 1998; Morrison, Ellis, & Quinlan, 1992). We did not control for the AoA factor in our experiments. Therefore, it is possible that any effects we attribute to the factor frequency might in fact be due to AoA. However, the discussion concerning the locus of the frequency effect in lexical access has been couched in terms of the factor frequency without controlling for AoA. It is not the purpose of this study to distinguish between these factors. Future experiments are required to disentangle which of these two factors is responsible for producing the effects discussed here.

whether the results obtained in the response association task generalise to a picture naming task.

EXPERIMENT 1A, 1B, AND 1C: PICTURE NAMING WITH NATIVE MANDARIN CHINESE

In Experiment 1a, a design was used that mirrored that of Jescheniak and Levelt (1994). Three sets of pictures with compound names in Mandarin Chinese were selected: (a) L(l) pictures, whose names are low frequency compound words (e.g., 熊猫, panda) and are composed of low frequency constituents (e.g., 熊, bear, and 猫, cat); (b) L(h) pictures (e.g., '天线', antenna), whose names are low frequency compound words but are composed of high frequency constituents (天, sky, and 线, line); and (c) H(h) pictures whose names are of high frequency (e.g., 电话, telephone) and are composed of high frequency constituents (e.g., 电, electricity, and 话, speech).

To illustrate how specific-word and morpheme frequencies were obtained consider the following example. The Chinese compound word (天线, antenna, /tian1 xian4/) is composed of two constituents – 天 (sky, /tian1/) and 线 (line, /xian4/). Whole-word frequency is simply the frequency of occurrence of the compound word '天线' (antenna), which happens to be 10 per 1.8 million (Institute of Language Teaching and Research, 1986). The cumulative (homophone) frequency count is the cumulative frequency of all the constituents having the same sound (regardless of their orthographic form), which in this case happens to be 5709 for /tian1/, and 6817 for /xian4/. By cumulative frequency we simply mean the sum of the frequencies of all the homophonic constituents. For example, 线 (line), 县 (town), 献 (bestow), 限 (constrain), etc., are all pronounced /xian4/. Therefore the cumulative frequency of /xian4/ is the sum of occurrences of all these cases. The reason for using cumulative frequency is, as laid out in the Introduction, that homophonic morphemes are represented by one lexeme node in one of the models (Levelt et al., 1999) that we are testing. The resulting morpheme frequency was obtained by averaging the cumulative frequencies of both constituents.

The rationale for choosing the average cumulative frequency, instead of, for example, first constituent frequency, is that the average constituent frequency is highly correlated with the individual constituent frequencies (the constituent frequencies of the first and second constituent were about equal, see Table 1 for details) and as such gives a good estimate of the role of constituent frequency in compound production. In addition, in all experiments in this paper we will also report regression analyses in which we

TABLE 1
Mean frequency distribution (range in parentheses) of the picture names in Experiment 1a: Mandarin Chinese. Along with the Mandarin examples we provide in parentheses the English translation and a literal English translation.

Picture Set	*Example*	*Compound frequency*	*1st Constituent frequency*	*2nd Constituent frequency*	*Constituents' mean frequency*
L(h)	天线 (antenna, sky-line)	10 (0–36)	3252 (803–13597)	4846 (758–34870)	4249
L(l)	熊猫 (panda, bear-cat)	8 (0–25)	509 (120–1678)	496 (13–2227)	502
H(h)	电话 (telephone, electricity- speech)	131 (31–307)	3703 (68–34870)	3727 (262–43820)	3715

specifically assess the contribution of first and second constituent frequency, among others, on naming latencies.[4]

The question we addressed in Experiment 1a is whether morpheme frequency or whole-word frequency predicts naming latencies for compound words. This is evaluated by comparing the three experimental conditions discussed above. The extent to which any differences between the experimental conditions are attributable to compound processing was tested in Experiment 1b. In Experiment 1b, we asked native English speakers to name in English the same set of pictures used with the Mandarin Chinese speakers. Given that the English translations of the Mandarin Chinese compound words are not themselves compounds (and barring any that are), Experiments 1a and 1b differ in one crucial respect: only in Experiment 1a is compound processing implicated; the other two variables – ease of picture recognition and name frequency – are the same in Experiment 1a and 1b. Thus, a comparison of the results in Experiments 1a and 1b will allow us to assess whether morpheme frequency affects naming performance over and above any effects of whole-word frequency and the ease with which a picture is recognised. Thus, for example, if we were to observe differences between picture sets L(l) and L(h) (i.e., a morpheme frequency effect) in Experiment 1a, but also in Experiment 1b, one could conclude that the morpheme frequency effect in Experiment 1a arises in picture recognition, not in compound production. This conclusion follows from the assumptions that name frequency between these conditions is matched (see below), and that picture recognition processes are comparable between Chinese and English speakers. Likewise, if we were to observe no morpheme frequency effect in Experiment 1a, a comparison with Experiment 1b would be useful to reveal whether such an effect is potentially masked by picture recognition processes.

Finally, in Experiment 1c we assessed the contribution of articulation factors to the performance in the experimental conditions of Experiment 1a. In Experiment 1c, participants perform a delayed naming task (Balota & Chumbley, 1985). In this task participants are presented with a word and name this word after a variable delay. If the delay is sufficiently long, it can be assumed that only factors related to articulatory execution can influence the word production process. Therefore, given that the same materials are used in Experiment 1a and 1c, Experiment 1c will provide a way to assess the contribution of articulatory factors in Experiment 1a.

---

[4] We acknowledge the potential importance of the role of semantic transparency in the representation of compounds (e.g., Libben et al., 2003 in word comprehension). However, the language production models under consideration here do not make assumptions about a role of semantic transparency in compound representation, and therefore, the predictions we derive from the models under consideration in this paper do not take into account the influence of this variable.

Methods

Participants. Twenty native Mandarin Chinese speakers from the Boston area took part in Experiment 1a. None of the participants had lived in the USA longer than 3 years. The data from one participant were discarded from the analyses because he tended to use a local dialect instead of standard Mandarin. Participants in Experiment 1b were 15 native English speakers from Harvard University, who were unfamiliar with Mandarin Chinese. Participants in Experiment 1c were 17 native Mandarin Chinese speakers from Beijing Normal University. All participants were paid or received course credits for participation.

Materials and design. Twenty-nine L(h) pictures were selected, each was paired with an L(l) picture matched for average whole-word frequency ($t < 1$) and with a H(h) picture matched for average morpheme frequency ($ts < 1$ for both of the two morphemes). All pictures were selected from the Snodgrass and Vanderwart (1980) picture set. The average morpheme frequencies of the L(h) and L(l) differed significantly from each other, $t(56) = 5.14$, $p < .001$. The average specific-word frequencies of L(h) pictures and L(l) pictures were lower than the whole-word frequency of H(h) pictures, (L(h) vs. H(h): $t(56) = 4.3$, $p < .001$; L(l) vs. H(h): $t(56) = 4.3$, $p < .001$. Mean frequencies are shown in Table 1 (see Appendix A for the complete list of stimuli). All the targets in the experiment were nominal bisyllabic compounds. It should be noted that the majority (over 70%) of Chinese words are bisyllabic compounds, and therefore having to produce only bisyllabic words in the experiment is unlikely to be perceived as odd by the participants.

We also included 37 pictures which served as fillers and as warm-up stimuli at the beginning of each block. The frequencies of the fillers varied over a large range. Responses to the filler pictures were not analysed. Participants were shown a total of 124 pictures (87 experimental pictures and 37 fillers). We also included a factor Presentation in the experiment. It has been shown in previous research that frequency does not interact with repeated presentation of stimuli (e.g., Jescheniak & Levelt, 1994). Each picture appeared once per block, and there were three blocks in the whole experiment. That is, each participant saw each picture three times in the actual experiment. Thus, the factor Presentation included three levels (first, second, third).

The same pictures were used in Experiment 1b. Most of the English names for the three picture sets are monomorphemic words, but there are also some compounds (5 in the L(h) condition, 4 in the L(l) and 4 in the H(h)). We considered the following types of frequency counts (from Francis & Kucera, 1982) for these English words: the whole-word frequency, the cumulative frequency (since some items have homophones; e.g., 'watch'), and morpheme frequencies for the few compounds.

The whole-word frequencies and the morpheme frequencies of the English names for the three sets of pictures were both distributed as the whole-word frequencies of the Mandarin Chinese names. Both types of frequencies for the English names of L(h) pictures (mean whole-word frequency: 11; mean cumulative frequency: 19) and L(l) pictures (mean whole-word frequency: 12; mean cumulative frequency: 19) are comparable (*t*s < 1), and are significantly lower than H(h) pictures (mean whole-word frequency: 46; mean cumulative frequency: 62), whole-word frequency: $t(56) = 4.5$, $p < .001$; cumulative frequency: $t(56) = 2.8$, $p < .01$. The morpheme frequency patterns of the few compounds in the three picture sets are also similar, and frequency counts in CELEX (Baayen, Piepenbrock, & van Rijn, 1993) were also checked and a similar pattern of frequency distribution was revealed. There were no significant differences in word length across sets (number of phonemes are 4.9, 4.9, 4.6 respectively, $F < 1$). That is, the English names for the L(h) and L(l) pictures are well matched across the three types of frequencies considered, and are significantly lower in frequency than the H(h) pictures.

In Experiment 1c, the names of the critical and filler pictures from Experiment 1a were used. All filler items appeared at the 700 or 1300 ms delay, while all critical items appeared at the 1000 ms delay (following Jescheniak & Levelt, 1994). Critical and filler items were composed into two blocks of trials with the same constraints as in Experiment 1a. An additional constraint was that there were no more than three consecutive trials with the same delay.

Procedure. In Experiment 1a and 1b, participants sat in front of an Apple Macintosh Powermac G3 computer with 17-in. monitor in a dimly lit room. The participants wore a headphone with attached microphone. The microphone was attached to a buttonbox equipped with a voicekey and provided measurements of vocal responses with a 1 ms resolution. The experimental software was Psyscope (Cohen, MacWhinney, & Flatt, 1993). In Experiment 1c, participants sat in front of PC (Pentium) and the experimental software was DMDX (Forster & Forster, 2003).

In both Experiments 1a and 1b there were three parts in the experiment. The first part consisted of familiarising the participants with the pictures that would appear later in the actual experiment. On each trial, participants first saw a fixation cross for 500 ms. Then the picture appeared for 2000 ms, and after 1000 ms its name appeared beneath the picture. Participants were required to name the picture aloud upon seeing its name. After a delay of 1000 ms, a new trial started.

The second part consisted of practicing the experimental task, while the third part consisted of the experiment proper. The second and third parts of the experiment were identical in trial structure. Participants first saw a fixation cross for 700 ms. Next, the picture appeared for 1500 ms or until the participant made a vocal response. Finally, there was a delay of 2000 ms, after which the next trial started.

Experiment 1c involved only a practice session of the experimental task, and the experiment proper. On each trial, participants first saw a fixation cross for 700 ms. Next, the stimulus word appeared for 600 ms after which the screen was blank for a variable delay of 700, 1000, or 1300 ms. Finally, a response cue ('*') appeared which remained on the screen for 1000 ms. After a delay of 2000 ms the next trial started. Each experiment lasted about 45 minutes.

Analyses. In Experiment 1a responses coded as errors included (a) names that were not designated as the target responses; (b) verbal dysfluencies (stuttering, hesitating, or not responding); and (c) voicekey recording errors. Responses longer than 2500 ms or shorter than 300 ms and responses that exceeded a participant's mean reaction time by more than 2.5 standard deviations were classified as outliers. Errors and outliers were excluded from the naming latency analyses. In 1c, responses longer than 2000 ms or shorter than 200 ms and that exceeded a participant's mean by 2.5 standard deviations were classified as outliers.

Separate ANOVAS were carried out on reaction times by subjects (F_1) and by items (F_2). There were two factors in the analyses: Picture Set (L(l) vs. L(h) vs. H(h) pictures and Presentation (first vs. second vs. third). Picture Set was considered a within-subject factor for the F_1 analyses and a between-subject factor for the F_2 analysis. The factor Presentation was treated as a within-subject variable in both F_1 and F_2 analyses. The same analyses were repeated with participant's error rate as the dependent measure. We also report the minF′ statistic, and the 95% confidence intervals (CI) on the differences of the means for the pair-wise comparisons in the participant analyses.

Results

Mean response latencies and mean errors for each Picture Set are shown in Table 2 for Experiment 1a, Table 3 for Experiment 1b and Table 4 for Experiment 1c. Figure 2 presents a histogram for different ranges of reaction times collapsed across repetitions for Experiment 1a.

TABLE 2

Picture naming latencies (ms) and errors rates for each of three presentations of the stimuli in Experiment 1a: Mandarin Chinese. Along with the Mandarin examples we provide the English translation and, in parentheses, a literal English translation.

Picture Set	*Chinese name*	*First*	*Second*	*Third*	*Mean*
L(h)	天线 (antenna, sky-line)	795 (5.4)	771 (3.4)	748 (3.1)	771 (4.0)
L(l)	熊猫 (panda, bear-cat)	790 (8.2)	766 (3.8)	757 (2.4)	771 (4.8)
H(h)	电话 (telephone, electricity- speech)	747 (2.7)	718 (2.7)	711 (1.8)	725 (2.4)

TABLE 3

Picture naming latencies (ms) and errors rates for each of three presentations of the stimuli in Experiment 1b: English.

Picture Set	*Chinese name*	*First*	*Second*	*Third*	*Mean*
L(h)	天线 (antenna, sky-line)	765 (6.4)	713 (4.6)	681 (3.4)	720 (4.8)
L(l)	熊猫 (panda, bear-cat)	747 (9.0)	701 (3.2)	679 (2.8)	709 (4.9)
H(h)	电话 (telephone, electricity- speech)	701 (3.0)	670 (2.5)	644 (3.0)	672 (2.8)

Experiment 1a. A total of 7.2% of the data points (3.7% erroneous responses and 3.5% outliers) were excluded from the analyses of naming latencies. ANOVAs on naming latencies revealed significant main effects of Picture Set, $F_1(2, 36) = 59.2$, $MSe = 668.8$, $p < .0001$; $F_2(2, 84) = 3.6$, $MSe = 18066.2$, $p < .04$; minF′ (2, 94) = 3.4, $p < .04$, and Presentation, $F_1(2, 36) = 16.9$, $MSe = 1309.2$, $p < .0001$; $F_2(2, 168) = 51.1$, $MSe = 706.6$, $p < .0001$; minF′ (2, 62) = 12.7, $p < .001$. There was no interaction between these two variables (both Fs < 1), suggesting that frequency effects remained constant across presentation repetitions (Caramazza et al., 2001; Jescheniak & Levelt, 1994; but see Griffin & Bock, 1998).

Further comparisons among the three Picture Sets were carried out. L(h) Pictures were named significantly slower than H(h) pictures, $F_1(1, 18) = 76.6$, $MSe = 779.5$, $p < .0001$, CI = ± 11 ms; $F_2(1, 56) = 5.0$, $MSe = 18810.1$, $p < .05$; minF′ (1, 63) = 4.6, $p < .04$, but did not differ from L(l) pictures (Fs < 1). The difference between L(l) pictures and H(h) pictures was also highly significant, $F_1(1, 18) = 78.3$, $MSe = 753.2$, CI = ± 10 ms, $p < .0001$; $F_2(1, 56) = 5.3$, $MSe = 19000.9$, $p < .05$; minF′ (1, 63) = 4.9, $p < .03$. The effects of Picture Set on error rate were significant only in the F_1 analysis, $F_1(2, 36) = 4.67$, $MSe = .002$, $p < .05$; $F_2(2, 84) = 2.4$, $MSe = .005$, $p = .1$; minF′ (2, 119) = 1.6, $p = .2$. Further comparisons show that the error rate for the L(h) picture set is similar to that for L(l) pictures ($F_1 < 1$), and is higher than that for H(h) pictures, $F_1(1, 18) = 4.4$, $MSe = .002$, $p < .05$, CI = ± .02.

Regression analyses were carried out to further compare the contribution of whole-word frequency and morpheme frequency on naming latencies of compound words. The following factors were treated as potential predictors: (a) the whole-word frequency of the compound word, (b) the frequency of each of its two constituents, (c) the mean frequency of the two constituents, (d) the minimum frequency of the two constituents, and (e) the maximum frequency of the two constituents. We used log-transformed frequencies as potential predictors of naming latencies. With a step-wise method (entry $p = .05$, removal $p = .10$), whole-word frequency was the only significant predictor entered in the model ($R^2 = 0.061$, $p = .021$). The inclusion of the various constituent frequency measures failed to contribute significantly to the explanatory power of the model (ps > .48).[5]

---

[5] Collinearity statistics on excluded variables revealed tolerance values of .95 for frequency of the first noun, .95 for frequency of the second noun, .92 for the mean frequency of the two constituents, .98 for the minimum frequency of the two constituents, and .90 for the maximum frequency of the two constituents. The largest condition index was 3.31 suggesting no problems of collinearity were present in the current data set.

TABLE 4
Picture naming latencies (ms) and errors rates for the stimuli in Experiment 1c.

Picture Set	*Chinese name*	*Mean*	*Error*
L(h)	天线 (antenna, sky-line)	400	6.5
L(l)	熊猫 (panda, bear-cat)	408	6.5
H(h)	电话 (telephone, electricity- speech)	400	8.1

Experiment 1b. Errors and outliers accounted for 4.2% and 6.5% of the responses, respectively. The results of ANOVAs indicated significant main effects of Picture Set, $F_1(2, 28) = 29.8$, $MSe = 962.3$, $p < .0001$; $F_2(2, 84) = 4.1$, $MSe = 15006.6$, $p < .05$; minF′ (2, 103) = 3.6, $p < .04$, and Presentation, $F_1(2, 28) = 14.0$, $MSe = 3943.5$, $p < .0001$; $F_2(2, 168) = 71.2$, $MSe = 1898.4$, $p < .0001$; minF′ (2, 40) = 11.7, $p < .001$. The interaction between the two variables was not significant, $F_1 < 1$; $F_2(4, 168) = 1.2$, $MSe = 1898.4$, $p = .39$; minF′ (4, 106) < 1, $p = .84$. Additional analyses revealed that naming latencies were faster for H(h) pictures than L(h) pictures, $F_1(1, 14) = 40.7$, $MSe = 1277.8$, $p < .0001$, CI = ± 16 ms; $F_2(1, 56) = 9.2$, $MSe = 12117.1$, $p < .01$; minF′ (1, 70) = 7.5, $p < .008$, and L(l) pictures, $F_1(1, 14) = 40.3$, $MSe = 778.5$, $p < .0001$, CI = ± 12 ms; $F_2(1, 56) = 4.2$, $MSe = 16910.6$, $p < .05$; minF′ (1, 65) = 3.8, $p < .06$. The difference between L(h) pictures and L(l) pictures was not significant, $F_1(1, 14) = 3.1$, $MSe = 830.7$, $p > .10$; $F_2 < 1$; minF′ (1, 68) < 1, $p > .5$. There were no differences in the error rates across

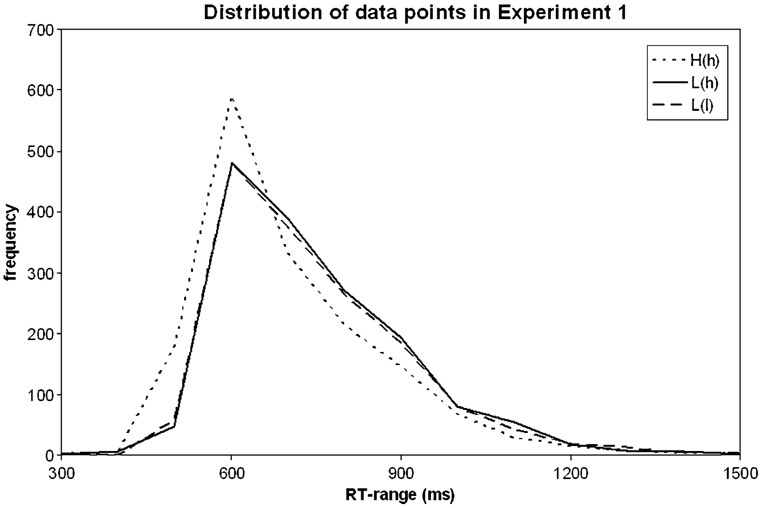

Figure 2. Distribution of mean naming latencies in Experiment 1a (Chinese) for the three types of compounds (L(h), L(l), and H(h)) across all three repetitions.

the three picture sets, $F_1(2, 28) = 2.3$, $MSe = .003$, $p > .1$; $F_2(2, 84) = 1.3$, $MSe = .010$, $p > .2$; $\text{minF}' (2, 105) < 1$, $p = .4$.[6]

Experiment 1c. A total of 7% of the data points (4.7% erroneous responses and 2.3% outliers) were excluded from the analyses of reaction times. The error analysis did not yield significant differences. In the RT analysis the factor Picture Set did not reach significance (*F*s < 1).

Discussion

The results of Experiment 1a revealed that low whole-word/high morpheme frequency (L(h)) items were produced more slowly than high whole-word/high morpheme frequency (H(h)) items and no faster than low whole-word/low morpheme frequency (L(l)) items. In other words, in Mandarin Chinese the main determinant of picture naming latencies appears to be whole-word frequency and not morpheme frequency (see also Chen & Chen, 2006).

Before turning to the interpretation of these results, there are several aspects of the data that deserve further discussion. First, the absence of a morpheme frequency effect in Experiment 1a could be attributed to a complex interaction between picture recognition times and compound retrieval times. That is, it could be argued that L(h) items were especially difficult to recognise and masked any advantages for the naming of these items due to the presence of high frequency morphemes. However, if such were the case, we would expect this factor to have contributed to the recognition times for the English speakers naming these items in Experiment 1b. In contrast to these predictions, no differences between L(l) and L(h) items were observed in this experiment. Therefore, it cannot be argued that the absence of the morpheme frequency effect in Experiment 1a is the result of slow recognition times for the L(h) items.

Likewise, it could be argued that the absence of a morpheme frequency effect could be attributed to a complex interaction between picture name articulation and compound word retrieval times. That is, it could be argued that the picture names in the L(h) condition were especially hard to articulate and masked any advantage for the naming of these items due to the presence of high frequency constituents. However, if such were the case, we would expect this factor to have contributed to the naming times of these items in

---

[6] As acknowledged earlier, in Experiment 1b there are several items for which the English translation is also a compound word. To investigate whether the inclusion of these compound words influenced the pattern of results obtained in Experiment 1b, we reanalysed the results excluding those words that are also compounds in English (see Appendix A for a list). Exactly the same results were obtained in Experiment 1a and 1b when we excluded the 13 items that translated into compounds in Experiment 1b.

Experiment 1c. In contrast to these predictions, no differences between the experimental conditions were observed.

In Experiment 2, native English participants produced compounds using the picture naming task of Experiment 1. Chen and Chen (2006) argued that one possible reason for the inconsistency between their results and those observed by studies using Dutch speakers is that Mandarin Chinese has a relative simple morphology while the morphology of Dutch is rather complex. If this were correct, then one would expect that compound production in English should pattern with Dutch, assuming that Dutch and English are comparable in terms of their morphological complexity. Thus, in Experiment 2, one would expect a morpheme frequency effect but no whole-word frequency effect.

One final point should be made. Compared with Mandarin Chinese, the English language presents with a much smaller set of compound words. The practical implication of this is that it was more difficult to find items for the experimental conditions in our experiment. Specifically, as it turned out, most picturable compounds have relatively low whole-word frequency, complicating the manipulation of this particular variable. Nevertheless, in order to test whether compound production is sensitive to whole-word frequency, we manipulated previously collected subjective familiarity ratings on the compounds as an estimate for whole-word frequency.

EXPERIMENTS 2A, 2B, AND 2C: PICTURE NAMING WITH NATIVE ENGLISH SPEAKERS

All critical items in this experiment were selected from a previously collected corpus of pictures with compound names. There are 164 pictures in this corpus. The pictures were collected from several sources (*Art Explosion*®, 1998; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997) and standardised such that they were all black and white drawings of roughly equal size. These pictures were subsequently rated by two groups of 20 participants. Ratings were collected according to the method described in Snodgrass and Vanderwart (1980). The first group rated the pictures on subjective familiarity, and name agreement norms were computed. The second group rated the pictures for name-image agreement and visual complexity.

The experimental conditions in Experiment 2a were similar to those in Experiment 1a, although the manipulation of whole-word frequency was now a manipulation of whole-word familiarity. This was because there was not sufficient variation in the written frequency counts obtained from CELEX to construct a manipulation of the compound's whole-word frequency. Therefore, instead of using written frequency counts, we decided to use the previously collected subjective familiarity ratings for the compound words.[7] We further discuss the implications of this design choice below. Three sets of pictures whose English names are compound words were selected: (a) L(l) pictures whose names are low familiarity compound words (e.g., bowtie) and are composed of low frequency constituents (e.g., 'bow' and 'tie'), (b) L(h) pictures whose names are low familiarity compound words (e.g., doghouse) and are composed of high frequency constituents (e.g., 'dog' and 'house'), and (c) H(h) pictures whose names are high familiarity compounds words (e.g., newspaper) and are composed of high frequency constituents (e.g., 'news' and 'paper').

To illustrate how familiarity and constituent frequency was obtained consider the following example. The English compound word 'newspaper' is composed of two constituents 'news' and 'paper'. Whole-word familiarity is simply the subjective familiarity rating of 'newspaper' that was collected previously, which happens to be 6.15 (range 1–7, 1 = unfamiliar, 7 = highly familiar). The morpheme frequency was calculated as in Experiment 1 (i.e., the average of the cumulative frequency of the compound's constituent homophones).

Like in Experiment 1, we again evaluated the performance obtained in Experiment 2a in terms of recognition and articulatory factors. In Experiment 2b, we assessed the contribution of recognition processes to the performance in the experimental conditions of Experiment 2a, by using the materials from Experiment 2a in a word-picture matching task. This picture recognition task is commonly used as a way to assess the degree to which the frequency effect arises as a result of input/semantic processes involved in picture naming (e.g., Almeida et al., 2007; Jescheniak & Levelt, 1994). In this task participants are presented with a word and are asked to make a yes/no decision on whether a subsequently presented picture matches the presented word. Importantly, it is assumed that this decision relies on processes involved in the recognition of pictures, and not in the retrieval of the name of the picture, and hence, if no effect of frequency is found in this task, an interpretation of the frequency effect in terms of input/semantic processes can be ruled out.

Importantly, the results from Experiment 2b can also be used to clarify the interpretation of a potential subjective familiarity effect in Experiment 2a. Although it has been shown that subjective familiarity ratings are accurate estimates of word frequency (Gernsbacher, 1984; Gordon, 1985), it has also been argued that an effect of the factor subjective familiarity is ambiguous between semantic and lexical processing (Balota, Pilotti, & Cortese, 2001; Snodgrass & Vanderwart, 1980). Experiment 2b reveals to

---

[7] A regression analysis on the compound pictures from this experiment (N = 42) with familiarity rating as a dependent variable revealed that the log transformed frequency of the first and second constituent were not significant predictors. Hence, the familiarity ratings collected for all compound pictures were not influenced by the compound's constituent frequency.

what extent a potential whole-word familiarity effect in Experiment 2a arises because of differences between the picture sets in terms of input/semantic factors. Thus, if an effect of subjective familiarity were found in Experiment 2a, and if the locus of this effect were in input/semantic processing, one would also expect this effect to arise in Experiment 2b. By contrast, if an effect of subjective familiarity was observed in Experiment 2a, but not in Experiment 2b, this would suggest a lexical locus of the subjective familiarity effect.

In Experiment 2c the same delayed naming task is used as in Experiment 1c.

Methods

Participants. Sixty Harvard University undergraduates participated in Experiment 2a, 2b, and 2c. There were an equal number of participants in each experiment and no participant took part in more than one experiment. All were native speakers of English. Upon completion of the experiment, participants were either paid or received course credit.

Materials and design. Roughly the same design as in Experiment 1 was used. Three sets of 14 pictures were chosen, each corresponding to the L(l), L(h), and H(h) conditions (see Appendix B). We controlled the following factors: number of syllables, $F(2, 39) = 1.92$, $p = .16$, word-length ($F < 1$), name-agreement, $F(2, 39) = 1.42$, $p = .25$, name-image agreement ($F < 1$), and visual complexity ($F < 1$). Names of pictures in the L(l) and L(h) sets did not differ on whole-word familiarity, $t(26) = 1.31$, $p = .20$, while their morpheme frequencies differed significantly, $t(26) = 3.72$, $p < .002$, for the average of both morpheme frequencies. Names of pictures in the L(h) and H(h) sets did not differ on morpheme frequency ($t < 1$), while their whole-word frequency/familiarities differed significantly, $t(26) = 8.14$, $p < .001$. Names of pictures in the L(l) and H(h) differed on whole-word familiarity, $t(26) = 5.28$, $p < .001$, and on morpheme frequency, $t(26) = 4.82$, $p < .001$. For an overview of the factors controlled in this experiment and their averages see Table 5.

In addition to the 42 critical pictures, a set of filler and practice pictures was selected. The filler pictures were selected from the Cycowicz et al. (1997) picture set. The selection of filler pictures was constrained by two criteria. First, the name of the picture could not be a compound word. Second, to avoid priming effects, the name of the picture could not be a constituent part of the critical compound words (e.g., apple – pineapple). In order to divert attention from the compound pictures in our experiment (in accordance with the appearance of compounds in the English language), we selected three times as many filler pictures as critical pictures, leading to a total of 126 filler pictures. The practice pictures were selected by taking a random sample of 20 pictures from the filler picture set.

TABLE 5
Average lexical statistics of items used in Experiment 2a, 2b, and 2c.

	Picture Set		
Statistic	*L(l)*	*L(h)*	*H(h)*
Familiarity	2.8	2.5	5.3
First morpheme frequency	38.2	152.8	243.4
Second morpheme frequency	55.2	292.3	228.9
Mean morpheme frequency	46.7	222.5	235.9
Name agreement	90.0	89.6	93.5
Name-image agreement	4.9	4.9	5.0
Visual complexity	4.0	3.8	3.7
Number of syllables	2.6	2.2	2.6
Number of letters	9.0	8.9	9.0

The critical and filler items were composed into two blocks. Equal numbers of items from each condition were assigned to each block. The order of blocks was counter-balanced, and half the number of participants received the reverse sequence of trials. The order of trials was pseudo-randomised with the following three constraints. First, a critical trial was always followed by at least one filler trial. Second, there were never any consecutive trials that shared the phonological onset of the target. Finally, there were never any consecutive trials on which there was a semantic relationship.

In Experiment 2b, we used the same critical and filler pictures. In order to elicit an equal number of 'yes' and 'no' responses from the participants, all items appeared twice. Each item was preceded once by its own name, evoking a yes-response, and once by another name evoking a no-response. The no-trials were composed as follows. The critical items were preceded by a word selected at random from the set of non-compound filler items. The filler items were preceded by a compound word on 42 trials, and preceded by a re-paired filler word on the remaining 84 trials. Care was also taken that there was no semantic or phonological overlap between the word and target picture on no-trials.

The experimental design of Experiment 2c was identical to Experiment 1c.

Procedure. The procedure in Experiment 2a and 2c was identical to that used in Experiment 1a and 1c, respectively. In Experiment 2b (word-picture matching), there were three parts in the experiment. In the first part they saw pictures appear on the screen for 1000 ms, and were asked to pay attention to the picture's name, but not say it aloud. After a delay of 1000 ms, a new trial started. The second part consisted of practicing the experimental task, while the third part consisted of the experiment proper. The second and third parts of the experiment were identical in trial structure. Participants first saw a fixation cross for 500 ms, followed by the presentation of a word. After the word had been on the screen for 1000 ms the picture appeared for 2000 ms. Participants pressed a button on the keyboard if the picture matched the word, and another button if they did not match. Finally, there was a delay of 1500 ms, after which the next trial appeared. Experiment 2b lasted about 40 minutes.

Analyses. Data trimming was carried out like in Experiment 1. In Experiment 2b, only the 'yes' trials were considered for analysis. Responses coded as errors included (a) false positives; (b) false negatives; and (c) no responses. The only factor in the analyses was Picture Set (L(l) vs. L(h) vs. H(h) pictures. Picture Set was considered a within-subject factor for the F_1 analyses and a between-item factor for the F_2 analysis.

Results

Table 6 presents the mean RTs and error percentages for each condition for Experiments 2a, 2b and 2c. Figure 3 presents a histogram for different ranges of reaction times for Experiments 2a.

Experiment 2a. A total of 10.5% of the data points (7.9% erroneous responses and 2.6% outliers) were excluded from the analyses of naming latencies. The error analysis did not yield significant differences. The RT analyses revealed a main effect of the factor Picture Set, $F_1(2, 38) = 8.4$, $MSe = 1972.5$, $p < .002$; $F_2(2, 39) = 3.4$, $MSe = 3634.9$, $p < .05$; $\text{minF}'(2, 66) = 2.4$, $p = .09$. Subsequent pair wise comparisons showed that the L(h) condition differed from the H(h) condition, $F_1(1, 19) = 18.1$, $MSe = 1746.3$, $p < .001$, $\text{CI} = \pm 28$ ms; $F_2(1, 26) = 6.6$, $MSe = 3342.4$, $p < .02$; $\text{minF}'(1, 41) = 4.8$, $p < .04$, but not from the L(l) condition, $F_1(1, 19) = 1.2$, $MSe = 2601.0$, $p = .29$, $\text{CI} = \pm$ 34 ms; $F_2 < 1$; $\text{minF}'(2, 42) < 1$, $p = .6$. In addition, the L(l) condition differed from the H(h) condition, $F_1(1, 19) = 9.5$, $MSe =$

TABLE 6
Mean picture naming latencies from Experiment 2a, picture recognition times for 'yes' trials from Experiment 2b, delayed naming latencies from Experiment 2c, and error percentages (in parentheses) for factor Picture Set.

	Picture Set		
Experiment	*L(l)*	*L(h)*	*H(h)*
Naming (Exp. 2a)	745 (7.5)	762 (13.9)	706 (10.0)
Matching (Exp 2b)	479 (9.3)	466 (5.4)	453 (5.7)
Articulation (Exp 2c)	354 (5.0)	361 (5.7)	356 (6.4)

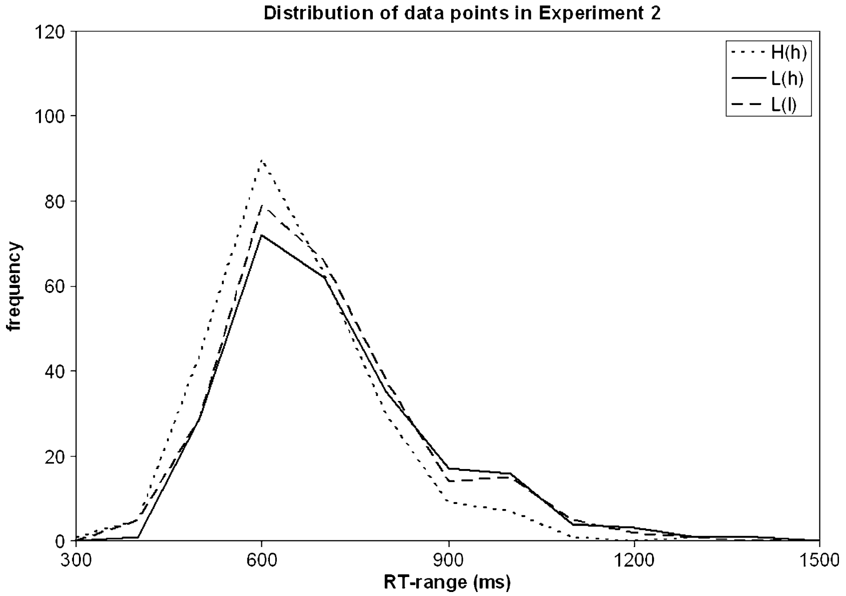

Figure 3. Distribution of mean naming latencies in Experiment 2a (English) for the three types of compounds (L(h), L(l), and H(h)).

1572.1, $p < .007$, CI = ± 26 ms; $F_2(1, 26) = 4.0$, $MSe = 3443.1$, $p < .06$; minF′(2, 42) = 2.8, $p = .1$. The error analyses revealed a main effect of the factor Picture Set by subjects but not by items, $F_1(2, 38) = 3.6$, $MSe = 1.2$, $p < .04$; $F_2(2, 39) < 1$; minF′ < 1.

Regression analyses were carried out to further compare the contribution of whole-word familiarity and morpheme frequency on naming latencies of compound words. We use log-transformed frequencies as potential predictors of naming latencies. The following factors were treated as potential predictors: (a) the subjective familiarity of the compound word, (b) the log frequency of each of its two constituents, (c) the mean log frequency of the two constituents, (d) the minimum log frequency of the two constituents, and (e) the maximum log frequency of the two constituents.

With a step-wise method (entry $p = .05$, removal $p = .10$), only the factor familiarity of the compound word significantly predicted reaction times (combined factors model $R^2 = 0.200$).[8] The inclusion of the other constituent

---

[8] Collinearity statistics on excluded variables revealed tolerance values of .93 for frequency of the first noun, .93 for the mean frequency of the two constituents, .93 for the minimum frequency of the two constituents, and .99 for the maximum frequency of the two constituents. The largest condition index was 4.88, suggesting no problems of collinearity were present in the current data set.

frequency measures failed to contribute significantly to the explanatory power of the model (ps > .08).

Experiment 2b. A total of 6.7% of the data points (4.3% erroneous responses and 2.4% outliers) were excluded from the analyses of reaction times. The error analysis did not yield significant differences. In the RT analysis the factor Picture Set did not reach significance, $F_1(2, 38) = 2.6$, $MSe = 1222.2$, $p = .09$, $F_2(2, 39) = 1.5$, $MSe = 2255.5$, $p = .23$; minF′(2, 72) < 1, $p = .4$. The error analysis also revealed no significant main effect of Picture Set, $F_1(2, 38) = 2.5$, $MSe = .8$, $p = .10$, $F_2(2, 39) = 1.6$, $MSe = 1.7$, $p = .22$; minF′(2, 74) < 1, $p = .4$.

Experiment 2c. A total of 5.7% of the data points (0.1% erroneous responses and 5.6% outliers) were excluded from the analyses of reaction times. In the RT analysis, the factor Picture Set did not reach significance (both Fs < 1). The error analysis also failed to reveal a significant main effect of Picture Set (both Fs < 1).

Discussion

The results reported in Experiment 2 revealed that the production of compounds in English, like in Mandarin Chinese, is sensitive to whole-word familiarity and not morpheme frequency. There are two aspects of the data that deserve further discussion. First, as we noted in the Discussion of Experiment 1, one possible explanation of the absence of a morpheme frequency effect is that this effect could be masked by slow picture recognition or response articulation times. That is, one could argue that the pictures of low whole-word familiarity – high morpheme frequency (L(h)) items were especially difficult to recognise or articulate compared with pictures from the low whole-word familiarity – low morpheme frequency L(l) items, thereby masking any advantage for the compound names of the L(h) items that would have accrued from having high frequency constituents. However, this explanation is again unlikely, given that these factors did not affect latencies in the recognition experiment (Experiment 2b) or the articulation experiment (Experiment 2c).

Second, the results of Experiment 2b suggest that the locus of the subjective familiarity effect observed in Experiment 2a is at the level of lexical processing. If the locus of the effect of subjective familiarity observed in Experiment 2a were at the semantic level, one would have expected this effect to arise in Experiment 2b in which differences between the picture sets in terms of input/semantic processing were assessed. Given that no effect was observed in Experiment 2b, it follows that the subjective familiarity effect observed in Experiment 2a can be explained in terms of lexical processing

(Balota, Pilotti, & Cortese, 2001; Snodgrass & Vanderwart, 1980). Note these results do not rule out a semantic locus of the subjective familiarity effect in general, but that our particular stimulus materials are not sensitive to semantic aspects of the subjective familiarity manipulation (see Alario, Costa, & Caramazza, 2002 for a similar argument).

A last point of discussion concerns the absence of the morpheme frequency effect and the possibility that this is due to a lack of power. Several aspects of the results make this possibility unlikely. First, as argued above, it is not the case that a potential morpheme frequency effect in the data was masked by recognition or articulatory factors, as demonstrated by the results from the four control experiments reported in this paper. Second, in both Experiment 1 and 2 the design was such that a morpheme frequency effect involved comparing the items of the L(h) condition with those in the L(l) condition, and a whole-word frequency effect involved comparing the same items of the L(h) and L(l) conditions with those in the H(h) condition. Thus, the absence of a morpheme frequency effect cannot be due to an ill-fated choice of the L(h) and L(l) items such that they are insensitive to reveal a frequency effect, given that those same items were sensitive enough to reveal a whole-word frequency effect. Third, the manipulation of morpheme frequency in the items of Experiment 1 and 2 was sufficiently strong to reveal a potential morpheme frequency effect. The ratio of the average high versus low cumulative morpheme frequency in Experiment 1 and 2 was 8.5, and 4.8, respectively. Previous experiments that have revealed significant frequency effects have used ratios of 12 (Jescheniak & Levelt, 1994), 6.5 (Griffin & Bock, 1998), or 4.8 and 3.1 (Bien et al., 2005, Experiment 1 and 2). Thus, the ratio used in the experiments reported here falls within the range of ratios used in previous studies. Finally, the absence of the morpheme frequency effect is replicated in the two experiments reported here, and is found in a series of independent studies on compound word (Chen & Chen, 2006), and homophone production (Bonin & Fayoll, 2002; Caramazza et al., 2001). Thus, it seems unlikely that a lack of power can explain the absence of the morpheme frequency effect in the current experiments.

GENERAL DISCUSSION

In two picture-naming experiments we manipulated the morpheme frequency and whole-word frequency of compounds. In the first experiment, it was found that native Mandarin Chinese speakers' picture naming latencies were determined by whole-word frequency, and not by morpheme frequency. Surprisingly, an identical pattern of results was obtained in the second experiment with native English speakers. These results were further confirmed by regression analyses, which showed that the sole reliable predictor of naming latencies was whole-word frequency. Four additional control experiments ruled out that recognition or articulatory processes were the cause of the observed differences between the sets of pictures. Thus, the results reported here indicate that in compound word production, whole-word and not morpheme frequency is the main determinant of picture naming latencies.

The results reported here have implications for the two models of lexical representation discussed in the Introduction. According to the two-stage model of lexical representation, compounds are stored in their decomposed form at the lexeme level, and the frequency variable affects the retrieval of lexeme representations (Levelt et al., 1999). On these assumptions it follows that compound word retrieval should be a function of the frequency of the compound's constituent morphemes, and hence, the main determinant of naming latencies should be a compound's morpheme and not its whole-word frequency. By contrast, according to a single-stage model of lexical representation (Caramazza, 1997), compounds are stored in their full-form at the lexical level. Assuming that frequency has an effect at this level, it follows that compound word retrieval should be a function of the compound's whole-word frequency. Thus, this model predicts that the main determinant in naming latencies should be the compound's whole-word and not its morpheme frequency. In line with the predictions of the latter model, the results from Experiment 1 and 2 revealed that naming latencies were sensitive to the compound's whole-word, but not its morpheme frequency.[9]

This interpretation of the results is further supported by other data. In a study by Chen and Chen (2006), Mandarin Chinese participants named compounds in the context of a response-association task. In line with the current findings, their results revealed that naming latencies were not sensitive to the compound's morpheme frequency. Thus, these results further support the interpretation of the current results in terms of a single stage model of lexical access.

However, there also exist results from studies that are problematic for the single stage model. Thus, in Roelofs (1996a) and Bien et al. (2005), native Dutch participants produced compound words in the context of a response-association task. In contrast to the results of Chen and Chen, and those of the current study, the results of these experiments revealed that compound word production was sensitive to morpheme frequency.

Why are the results obtained in the current study and those obtained in the study of Chen and Chen (2006) different from those obtained by Roelofs

---

[9] These results do not rule out the class of models with a lemma/lexeme distinction as a whole. If such models were to assume that compounds are stored in their full form at the lexeme level, or if they were to assume that the frequency variable affects lemma retrieval (with decomposed representations at the lexeme level), they would predict the results obtained here.

(1996a) and Bien et al. (2005)? As discussed by Chen and Chen (2006), one possible explanation for this discrepancy could be in terms of *cross-linguistic differences*. They argued that morphological representations perhaps do not play a role in languages with a simple morphology like Mandarin Chinese, while they do play a role in languages with complex morphology like Dutch – thereby explaining the absence of the morpheme frequency effect in Mandarin Chinese and the presence of the effect in Dutch. In the present study this possibility was further investigated by comparing compound word production in Mandarin Chinese to English. It was assumed that Dutch and English have comparable morphology, and so on the basis of the cross-linguistic differences proposed by Chen and Chen, one would predict that compound word production in English should pattern with Dutch and not with Mandarin Chinese. In contrast to this prediction, the present results revealed that English patterns with Mandarin Chinese, and not with Dutch. Thus, on the basis of these results one could conclude that it is unlikely that cross-linguistic differences can explain the discrepancy in the results.

However, an obvious objection to this conclusion would be to argue that Dutch and English do not have comparable morphology, and hence, it is not surprising that the results obtained in English do not pattern with those obtained in Dutch. There is evidence to suggest that the morphological system of the two languages might be different (e.g., Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005). In this case, the results obtained here would not be inconsistent with the possibility that cross-linguistic differences impact the representation of compounds. One would then have to conclude that compound representations are comparable in Mandarin Chinese and English but different from Dutch. This seems highly unlikely but it should receive further investigation.

Another possible explanation of the discrepancy in the results observed in the present study and those previously observed (Bien et al., 2005; Roelofs, 1996a) could be in terms of *task-differences*. The earlier studies all used a response-association task, while in the current study compound word production took place on the basis of a picture naming task. One possibility is that the picture naming task is sensitive to other aspects of compound word production than the response-association task. If it is assumed that the picture-naming task is primarily sensitive to whole-word and not morpheme retrieval, then one would expect naming latencies in the picture naming task to reveal whole-word but not morpheme frequency effects. If it is assumed that the response-association task is more sensitive to morpheme retrieval than to whole-word retrieval processes, then one would expect naming latencies in this task to reveal morpheme but not whole-word frequency effects. Such differences in the sensitivity of tasks to whole-word or morpheme retrieval might stem from differences in the way that the two tasks rely on conceptually driven lexical access. Thus, while it is generally accepted that the picture naming task reveals conceptually mediated word production (e.g., Potter & Faulconer, 1975), evidence concerning this issue in the response association task is sparse (e.g., Roelofs, 1996b).

The discussion above illustrates that the impact of cross-linguistic and task-differences on compound word production are not well understood. At present, this state of affairs limits a clear interpretation of the data on compound word production in terms of a model of lexical access. One possible avenue for future research would be to consider an experiment in which compound word production takes place with Dutch speakers in a picture naming task. If such an experiment were to reveal that picture naming latencies were sensitive to morpheme and not whole-word frequency effects, this would suggest that cross-linguistic rather than task-differences have an impact on compound-word production. By contrast, if this Dutch picture naming experiment were to reveal whole-word and not morpheme frequency effects, this could suggest that task, and not cross-linguistic differences impact compound word production. Thus, a study of this sort would be able to resolve some of the issues discussed here.

To conclude, we have found that the effective frequency in determining picture naming latencies of compound words is the frequency of the compound as a whole, and not the frequency of the constituent morphemes. These results are consistent with a single-stage model of lexical access (e.g., Caramazza, 1997), and inconsistent with a two-stage model that assumes a decomposed representation of compounds and a lexeme locus of the frequency effect (e.g., Levelt et al., 1999). Although this interpretation of the results in terms of a single-stage model is supported by previously reported results in Mandarin Chinese (Chen & Chen, 2006), it is not supported by results that differ from the current results in terms of the target language and task (Bien et al., 2005; Roelofs, 1996a). Future studies examining the production of compounds with Dutch participants using a picture naming task will be fruitful in advancing the research discussed here.

REFERENCES

Alario, F.-X., Costa, A., & Caramazza, A. (2002). Hedging one's bets too much? A reply to Levelt (2002). *Language and Cognitive Processes*, *17*, 673–682.

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect: When recognizing is not enough. *Psychonomic Bulletin and Review*, *14*, 1177–1182.

Art Explosion (1998). NOVA Development [CD-ROM]. Calabasas, CA.

Baayen, R. H., Piepenbrock, R., & van Rijn, A. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89–106.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition*, *29*, 639–647.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long term semantic priming: A computational account and expirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1059–1082.

Bien, H., Levelt, W. J. M., & Baayen, H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences*, *102*, 17876–17881.

Bonin, P., & Fayoll, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, *14*, 289–313.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production: Vol. II. Development, writing and other language processes* (pp. 257–294). London: Academic Press.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., Miozzo, M., Costa, A., & Bi, Y. (2001). The whole-word frequency effect: Implications for the representations of homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1430–1450.

Carroll, J. B., & White, M. N. (1973b). Word frequency and age of acquisition as determiners of picture naming latencies. *Quarterly Journal of Experimental Psychology*, *25*, 85–95.

Chen, T-M., & Chen, J-Y. (2006). Morphological encoding in the production of compound words in Mandarin Chinese. *Journal of Memory and Language*, *54*, 491–514.

Cohen, J. D., MacWhinney, B., & Flatt, M. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments and Computers*, *25*, 257–271.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237.

Damian, M. F., & Als, L. C. (2005). Long lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1372–1384.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, *81*, 555–567.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*, 313–349.

Dohmes, P., Bölte, J., & Zwitserlood, P. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language*, *90*, 203–212.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 515–523.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, *35*(1), 116–124.

Francis, N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Miffin Company.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). London: Academic Press.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language*, *24*, 631–645.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing and levels in spoken word production. *Journal of Memory and Language*, *38*, 313–338.

Gumnior, H., Bölte, J., & Zwitserlood, P. (2006). A chatterbox is a box: Morphology in German word production. *Language and Cognitive Processes*, *21*, 920–944.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464–482.

Institute of Language Teaching and Research, (1986). *Modern Chinese Frequency Dictionary (Xiandai Hanyu Pinlv Cidian)*. Beijing, China: Beijing Language Institute Press.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.

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 decision, and eye fixations. *British Journal of Psychology*, *94*, 223–244.

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, H. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, *20*, 43–73.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, *23*, 39–66.

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory and Language*, *31*, 333–348.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

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.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: The phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object recognition. *Memory and Cognition*, *50*, 705–714.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.

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

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature*, *253*, 437–438.

Roelofs, A. (1996a). Morpheme frequency in speech production: Testing WEAVER. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology*. Dordrecht, the Netherlands: Kluwer Academic Press.

Roelofs, A. (1996b). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854–876.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin and Review*, *13*, 33–37.

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.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 761–775.

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

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology*, *44A*, 723–761.

Wingfield, A. (1967). Perceptual and response hierarchies in object identification. *Acta Psychologica*, *26*, 216–226.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, *81*, 226–234.

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

Zwitserlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production: Evidence from picture naming. *Language and Cognitive Processes*, *15*, 563–591.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2002). Where and how morphologically complex words interplay with naming pictures. *Brain and Language*, *81*, 358–367.

APPENDIX A

Critical pictures in Experiment 1a (Chinese), 1b (English), and 1c (Chinese).

Condition		
L(h)	*L(l)*	*H(h)*
斑马 zebra	皇冠 crown	嘴唇 lips
阳台 balcony	鸵鸟 ostrich	蜡烛 candle
土豆 potato	芹菜 celery	苍蝇 fly
樱桃 cherry	恐龙 dinosaur	耳朵 ear
羽毛 feather	熨斗 iron	窗户 window
山羊 goat	螃蟹 crab	玉米 corn
信封 envelope	帐篷 tent	冰箱 refrigerator
火山 volcano	烟斗 pipe	太阳 sun
花瓶 vase	台灯 lamp	老虎 tiger
水壶 kettle	秋千 swing	手套 glove
手表 watch	钢琴 piano	苹果 apple
沙发 couch	轮胎 tire	水池 basin
气球 balloon	盔甲 armor	青蛙 frog
蛋糕 cake	钮扣 button	手枪 gun
开关 switch	蜗牛 snail	眼镜 glasses
脚趾 toe	辣椒 pepper	尾巴 tail
香肠 sausage	扫帚 broom	面包 bread
天平 scale	黄瓜 cucumber	月亮 moon
鲸鱼 whale	楼梯 stairs	电视 television
毛衣 sweater	抽屉 drawer	眼睛 eye
箭头 arrow	萝卜 carrot	电话 telephone
天线 antenna	松鼠 squirrel	火箭 rocket
公鸡 rooster	漏斗 funnel	衣服 coat
* 花生 peanut	* 磁盘 floppy-disk	* 汽车 bus
* 足球 soccer-ball	* 菠萝 pineapple	* 灯泡 light-bulb
* 毛笔 paintbrush	* 牙刷 toothbrush	* 警察 policeman
* 草莓 strawberry	* 熊猫 panda	* 大脑 brain
* 和尚 monk	* 孔雀 peacock	* 报纸 newspaper
* 金鱼 gold-fish	* 灯笼 lantern	* 飞机 airplane

*Excluded in the additional analyses of Experiment 1a and 1b: ‘Comparison of Experiments 1a and 1b for monomorphemic words in English’.

APPENDIX B

Critical pictures in Experiment 2a, 2b, and 2c.

Condition		
L(l)	*L(h)*	*H(h)*
Roller-skate	Tape-measure	Mail-box
Beehive	Lady-bug	Paper-clip
Corkscrew	Rainbow	Cup-cake
Totem-pole	Dartboard	Credit card
Marshmallow	Windmill	Basket-ball
Mousetrap	Wheel-chair	Headphones
Bowtie	Christmas-tree	Keyboard
Pineapple	Beach-ball	Baseball
Wheelbarrow	Cowboy	Toilet paper
Motor-cycle	Sea-horse	Light bulb
Lipstick	Hour-glass	Newspaper
Butterfly	Bird-house	Briefcase
Ice-skate	Dog-house	Water-melon
Sail-boat	Lighthouse	Fire-place