Mem Cogn
DOI 10.3758/s13421-014-0396-z

Task influences on the production and comprehension of compound words

Niels Janssen · Petra E. Pajtas · Alfonso Caramazza

Abstract Studies of compound word processing have revealed effects of the compound's constituents in a wide variety of word production and comprehension tasks. Surprisingly, effects of the compound's constituents were not found in a recent word production study using the picture-naming task. Here, we examined whether these contrasting constituent effects reflect methodological differences or whether they reflect the influence of task-specific characteristics on compound processing. A large set of compounds was used to elicit picture-naming latencies in Experiment 1 and visual lexical decision latencies in Experiment 2. Regression analyses revealed surface frequency effects in picture naming and lexical decision and constituent effects only in lexical decision. These results rule out methodological differences as the basis for the contrasting constituent effects observed between the various tasks and imply that constituent effects are best understood as arising out of task-specific characteristics.

Keywords Language production · Language comprehension · Morphology · Compounds

The study of familiar compound words (e.g., *doghouse*) has been used in the fields of word comprehension and production to gain insight into the role of morphology in the mental lexicon. The general finding is that effects of the compound's constituents have been observed in a variety of word comprehension tasks, such as lexical decision and word reading (e.g., Andrews, 1986; De Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002; Duñabeitia, Perea, & Carreiras, 2007; Juhasz, Starr, Inhoff, & Placke, 2003; Kuperman, Schreuder, Bertram, & Baayen, 2009; Libben, Gibson, Yoon, & Sandra, 2003; Pollatsek, Hyönä, & Bertram, 2000), and in word production tasks, such as response association and picture–word interference (e.g., Bien, Levelt, & Baayen, 2005; Dohmes, Zwitserlood, & Bölte, 2004; Gumnior, Bölte, & Zwitserlood, 2006; Koester & Schiller, 2008; Roelofs, 1996; Zwitserlood, Bölte, & Dohmes, 2000). In light of these data, a rather surprising result was recently reported by Janssen, Bi, & Caramazza (2008). That study found that compound picture naming did not yield the constituent effects found in other tasks. Here, we examined whether these contrasting data reflect methodological differences between the studies or whether they reflect task-specific influences on the processing of compounds. To address this issue, we investigated compound word processing in picture-naming and lexical decision tasks.

The majority of studies on compound word processing have been conducted in the domain of visual word recognition. To examine the role of the compound's constituents in lexical processing, these studies rely on a manipulation of measures associated with the compound's whole word and constituents. Many studies have revealed effects of the lexical frequency of the compound's constituents and its surface (whole-word) form. For example, Juhasz et al. (2003) asked English participants to perform naming, lexical decision, and reading in sentence context tasks on visually presented compound words while their eye movements and reaction times (RTs) were registered. The frequency of the first and second constituents of the compounds was varied factorially, while the compound's surface frequency was held constant. The results revealed effects of the second constituent frequency on both eye fixations and latencies in all three tasks. Similarly, Duñabeita et al. (2007) found that lexical decision times were affected by

N. Janssen (✉)
Cognitive Neuroscience and Psycholinguistics Lab,
Universidad de La Laguna, La Laguna 38205, Spain
e-mail: njanssen@ull.es

P. E. Pajtas · A. Caramazza
Harvard University, 33 Kirkland Street, Cambridge,
MA 02138, USA

A. Caramazza
Center for Mind/Brain Sciences, University of Trento, Trento, Italy

Published online: 02 April 2014

the frequency of the second constituent in both head-final Spanish and head-initial Basque compound words. Finally, Pollatsek et al. (2000) revealed that both surface frequency and second constituent frequency affected eye fixations in a reading in sentence context task with Finnish compound words.[1]

The impact of morphology on lexical processing has also been examined using other measures than lexical frequency. For example, Kuperman et al. (2009) conducted a large scale study involving 2,500 compounds in which Dutch participants performed a lexical decision task while their eye movements and decision latencies were registered. Besides examining the impact of the compound's surface and constituent frequency on the dependent measures, they also considered variables related to the compound's constituent family size.

The variable constituent family size of a compound is defined as the cardinality of the set of compounds that share a constituent with a given compound in a given position. Thus, for example, for the following set of compounds {windmill, windshield, windbag, windhose}, the left constituent family size would be 4. In a regression analysis that relied on a mixed effect technique, Kuperman et al. (2009) reported that lexical decision latencies were sensitive to the compound's surface frequency, to constituent frequency, and to the compounds' constituent family size measures (see also De Jong et al., 2002; Juhasz & Berkowitz, 2011).

Although different models of complex word recognition have been proposed to account for these results (e.g., Caramazza Laudanna, & Romani, 1988; Seidenberg & Gonnerman, 2000; Taft, 2003; see Diependaele, Grainger, & Sandra, 2012, for a recent review), a shared assumption in all these models is that in the course of compound word recognition, the compound's constituents are activated and influence word comprehension.

Studies from the field of word production have also found evidence for a role of the compound's constituents in processing (Bien et al., 2005; Roelofs, 1996). For example, Bien et al. (2005) found constituent effects in compound word production using the response association task. In this task, participants first learn to associate a number of cues (i.e., other words, spatial positions on the computer screen) with visually or auditorily presented compound words. In the actual experiment, the cues are then used to elicit overt production of the associated compound word. Bien et al. found that production latencies were sensitive to various measures (i.e., frequency, family size) associated with the compound's constituents.

Other word production studies using the picture–word interference task have yielded comparable results (Dohmes et al., 2004; Gumnior et al., 2006; Koester & Schiller, 2008; Zwitserlood, Bölte & Dohmes, 2000, 2002). For example, Zwitserlood et al. (2000) reported shorter naming latencies to pictures with monomorphemic names (e.g., dog) when they were preceded by a visually presented compound word prime that contained the picture name (e.g., *doghouse*). On the basis of these data, the authors argued in favor of a model that assumes that a compound's constituent morphemes play a role in word production.

The overall conclusion that emerges from this brief review of word comprehension and production studies is that compound processing yields constituent effects. Given this evidence, it is surprising that rather different results were recently obtained by Janssen et al. (2008) using a picture-naming task. Janssen et al. (2008) asked Mandarin Chinese and English speakers to name pictures with compound names. The compound's surface frequency and constituent frequencies were orthogonally manipulated in a factorial design. The results revealed that in both Chinese and English, naming latencies were determined by the compound's surface frequency, but not by its constituent frequencies. Four additional control experiments ruled out the possibility that these results were artifacts related to picture recognition or articulatory processes. The absence of constituent effects in compound picture naming are surprising, considering the success of the task in elucidating how other components, such as semantics and phonology, play a role in the production of words (e.g., Glaser, 1992). In other words, the data reported by Janssen et al. (2008) raise the question of why the task would be specifically insensitive to constituent effects in compound word production.

Thus, whereas effects of the compound's constituent morphemes have been repeatedly reported in many studies using a variety of tasks, they were not found in the picture-naming study by Janssen et al. (2008). This contrast suggests that task-specific characteristics may have an important influence on how compound words are processed. However, before such a conclusion can be accepted, two issues need to be addressed. First, the empirical evidence for the absence of constituent effects in the picture-naming task is rather weak; it relies on a single study. Consequently, establishing the reliability of the picture-naming data reported by Janssen et al. (2008) is crucial. Second, there are many methodological differences between the previously reported studies and that of Janssen et al. (2008). The studies used different languages, materials, and analyses, and all of these might impact the results reported for compound word processing. In order to satisfactorily address the difference in findings, it is important that we examine critically whether the contrasting results can be reduced to methodological differences.

The two experiments reported below were designed to address these issues. In Experiment 1, we attempted to establish the reliability of the picture-naming results. The

[1] There is disagreement in the word recognition literature about the presence of first-constituent effects (cf. Juhasz et al., 2003; Taft & Forster, 1976). Here, we simply wish to highlight that effects of the constituents are found.

experimental design differed from the factorial design used by Janssen et al. (2008). Native English speakers named a large set of pictures ($N = 150$) with compound names whose surface and constituent measures varied (see below for details). Each participant named each picture twice—once before and once after a picture familiarization stage. This was done to address the possibility that the results observed by Janssen et al. (2008) were due to the familiarization procedure. In Experiment 2, we examined visual lexical decision latencies from the English Lexicon Project (Balota et al., 2007) for a subset of the compounds used in Experiment 1. We focused on the lexical decision task since this task has a long history in revealing robust constituent effects. In both experiments, we relied on the same analytical techniques to determine the presence of constituent effects.

Thus, in Experiments 1 and 2, comparable sets of compounds from the same language were analyzed in the same way. If the contrasting constituent effects observed in the literature arose as a consequence of differences in languages, stimulus sets, or analyses, we would not expect contrasting constituent effects between Experiments 1 and 2.

Experiment 1: Picture naming

Method

Participants

Twenty-eight native English speakers who were students at Harvard University participated in the experiment. All participants received course credit or compensation ($10 per hour).

Materials and design

All stimuli were selected from the database described by Janssen, Pajtas, and Caramazza (2011). One hundred and fifty pictures with compound names were used. They were black and white line drawings that fit on a white square of 300 × 300 pixels. All compounds had two major constituents.

The 150 pictures were assembled into two blocks of 150 trials each. The manipulation of block served to examine the effect of familiarization, where naming in block 1 occurred before familiarization and naming in block 2 occurred after familiarization. Each picture appeared once per block. The order of trials was pseudorandomized such that, on successive trials, picture names were semantically and phonologically unrelated.

In the regression analyses reported below, we distinguished between critical and control variables. The critical variables were related to the compounds' surface and constituent measures. For each compound, we computed surface and constituent (left and right) measures using CELEX data (Baayen, Piepenbrock, & van Rijn, 1993). Specifically, the compound's *surface frequency* was the lemma frequency (stem plus inflectional variants) of the whole compound word. Its left *lemma frequency* was the lemma frequency of the left constituent, and its left *cumulative frequency* was the left constituent's lemma frequency plus the sum of all the frequencies of complex words in which the constituent appeared, taking into account homophonic variants of the constituent (i.e., homophone frequency).

Other measures were based on the compound's *constituent family* (De Jong et al., 2002; Schreuder & Baayen, 1997), which is the set of compounds (or other derived words) that share a constituent (e.g., {windmill, wind chill, windbag, windshield, windy}). The left *constituent family size* was the cardinality of the compound's left constituent family; its left *positional frequency* was the sum of the lemma frequencies of the family members, and its left *positional entropy* was a measure of the distribution of the relative frequencies of the family members using Shannon's entropy (Shannon, 2001). The compound's left *complement frequency* referred to the sum of the frequencies of all complex, noncompound words in which the constituent appeared, and Shannon's entropy of this set was called its left *derivational entropy*. Right constituent measures were computed accordingly. Table 1 presents a summary of the distribution of these variables. Finally, the variable *block* denoted the number of times a picture was named (i.e., before and after familiarization).

In addition, we included a number of control variables that are known to affect latencies in psycholinguistic experiments in general, as well as the picture-naming task in particular. First, we considered the impact of the variable trial (the ordinal position of an item in the experiment). In addition, we examined the impact of subjective familiarity, name–image agreement, and visual complexity ratings that were obtained for these pictures (see Janssen et al., 2011, for details). The latter variables are typically associated with picture recognition processes (e.g., Snodgrass & Vanderwart, 1980). We also considered the variable compound word length in phonemes (Bertram & Hyönä, 2003; Juhasz, 2008). Finally, the impact of three articulatory factors (i.e., voicing, plosiveness, fricativeness) was taken into account, given the observation that such factors affect latencies in word production tasks (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004).

Procedure

The experiment was presented using DMDX (Forster & Forster, 2003). In the first block, participants named a picture and, subsequently, saw its name printed on the screen. Participants were told to use this name for the particular picture. This was considered the familiarization stage. On each trial, a fixation cross was presented (700 ms), followed by a blank screen (200 ms), the presentation of the picture (2,000 ms), the picture's name (1,000 ms), and a blank screen (1,000 ms). In the second block participants were instructed to name the

Table 1 Distribution of the variables associated with the compound's left and right constituents in Experiments 1 and 2

	Variables	Left Constituent		Right Constituent	
		Median	Range	Median	Range
Experiment 1	Lemma frequency	894	0–14,935	621	4–29,231
	Cumulative frequency	1,675	0–111,823	1,575	0–82,033
	Family size	7	1–63	13	1–426
	Positional frequency	68	0–3,038	77	0–6,600
	Positional entropy	309,396	0–579,678	603,204	0–2,698,981
	Complement frequency	29	0–5984	34	0–7,182
	Derivational entropy	258,307	0–555,677	278,545	0–642,252
Experiment 2	Lemma frequency	1,102	4–12,983	586	4–29,231
	Cumulative frequency	1,840	14–111,823	1,716	6–82,033
	Family size	10	1–63	16	1–426
	Positional frequency	130	1–3,038	114	0–6,600
	Positional entropy	198,000	0–500,174	310,730	0–2,698,981
	Complement frequency	49	0–4,510	45	0–7,182
	Derivational entropy	280,804	0–555,677	271,775	0–597,187

pictures again, and were told the picture's name would not be presented, and to use the name presented in the previous part. In this block, each trial began with a fixation cross (700 ms), followed by a blank screen (200 ms), the picture (2,000 ms), and another blank screen (1,000 ms).

Analysis and results

Trials on which participants produced incorrect responses, as well as trials on which the naming latency was less than 300 ms or greater than 2,500 ms, were discarded (1,940 trials out of 8,400, or 23 %). The resulting set of 6,460 data points was analyzed using linear mixed-effects methodology (Baayen, 2008). Continuous predictors, except for RT, were log-transformed to reduce skewness.

Statistical considerations

The full model that included all the critical and control variables described above suffered from high multicollinearity (condition index = 48). High collinearity in regression models is problematic for two reasons. First, high collinearity undermines a clear interpretation of the potential effects of predictors in the model, and second, high collinearity can lead to problems in the stability of the model-fitting procedures. We addressed the issue of collinearity with two measures.

First, all continuous variables were centered on their sample means. Centering of variables both improves the interpretability of variables in the model and reduces collinearity when predictors participate in interactions with other predictors. Second, correlated variables were residualized. Residualizing in regression designs has the same effect as controlling for a variable in a basic factorial design (Baayen, 2010). Given two correlated variables A and B, the residuals from a linear regression of A onto B will be uncorrelated with B and highly correlated with A. Using this procedure, we residualized variables when their Pearson correlation coefficient was higher than .15. All residualized variables were uncorrelated with the target variable and correlated highly with the original variable (r>.96).

In addition, model-fitting procedures often have problems with regression designs containing many variables. Under such circumstances, the model-fitting procedure may fail to converge. In our case, the full model that had the appropriate fixed-effect and random-effect structure failed to converge. Under such circumstances further reduction of the complexity of the statistical model is recommended (Barr, Levy, Scheepers, & Tily, 2013). Specifically, we distinguished in the model between those variables related to the compound constituent frequency and those related to the compound constituent family size. We opted to investigate the impact of these two sets of variables in analyses of two distinct statistical models. Furthermore, because there was a high degree of correlation between the constituent measures related to family size, and because the main focus of the study was not on distinguishing between these individual measures, a principal component analysis was conducted on all constituent family size variables. We extracted only the first component that captured the maximum amount of variance. We refer to the first principal components corresponding to the compound's left and right constituent family size measures as the left and right principal components. The left and right principal components captured 70 % and 72 % of variance, respectively.

In sum, variables were centered and residualized in order to reduce problems of collinearity. In addition, problems with convergence were addressed by implementing a principal component analysis and by analyzing the data with two different statistical models. The correlation matrix that lists the relevant predictors for the analyses reported below is presented in Table 2. Note that this matrix reports correlations before and after the residualization procedures.

Analysis 1: Basic frequency measures

The first analysis focused on the effects of basic frequency measures. The statistical model included the critical variables surface frequency, left and right constituent lemma frequency, all control variables (see above), block (before and after familiarization), and the interaction of block with all critical variables. The random-effect structure included by-participant and by-item random intercepts, by-participant random slopes for the three frequency variables, and by-item random slopes for the variable block. The by-participant random-effect structure did not include a parameter that estimated the correlation between the random variables. More complex random-effect structures led to models that failed to converge.[2]

The results from the mixed effect analysis of the full model is presented in the top-panel of Table 3. The heading "Full model" lists the beta-coefficients and *t*-values corresponding to an analysis in which all predictors are present simultaneously in the model. These results reveal important information about the sign of the various effects. The heading "Model comparisons" provides the significance value of the predictor. The model comparisons were performed using the likelihood ratio test, where models with and without predictors were compared while the random-effects structure was kept unchanged (Barr et al., 2013). A graphical presentation of the results from this analysis is presented in Fig. 1.

The results revealed an effect of block (Fig. 1a), where latencies became faster in the second block. There was also an effect of surface frequency in block 1 (Fig. 1b), where latencies became shorter with increasing frequency values. The effect of surface frequency interacted with block. In addition, there was an effect of trial (Fig. 1c), where mean naming latencies became longer along the course of the experiment.[3] In addition, there were effects of familiarity (Fig. 1d), name–image agreement (Fig. 1e), and visual complexity (Fig. 1f). Crucially, there were no effects of the compound's left and right constituent lemma frequencies.[4]

We further examined the interaction of block and surface frequency by inspecting the surface frequency effect separately in block 1 and 2. The statistical models in these analyses were identical to the model above, except that the predictor block and all its interactions were removed. There was an effect of surface frequency in block 1, $t(2699) = -2.65$, $\chi^2(1) = 7.39$, $p < .0066$, and in block 2, albeit smaller in size, $t(3741) = -2.49$, $\chi^2(1) = 6.51$, $p < .0108$.

Analysis 2: Constituent family size measures

In this analysis, we considered the impact of the variables related to the compound's constituent family size. Critical predictors in this statistical model were the compound's surface frequency and its left and right principal components. As before, we also included all control variables, block, and the interaction with block and the critical variables. The random-effects structure was similar to that of the first analysis. There were by-participant and by-item random intercepts, by-participant random slopes for surface frequency and the two principal components, and by-item random slopes for the variable block. More complex random-effect structures failed to converge (see footnote 2).

The results from the mixed-effect analysis of the full model is presented in the bottom panel of Table 3. The results paralleled those found in the first analysis. There was an effect of block, an effect of surface frequency that interacted with block, an effect of trial, and effects of familiarity, name–image agreement, and visual complexity. As in the first analysis, there were no effects of the left and right principal components related to the compound's constituent family size measures. The interaction of surface frequency with block was further examined. As in the first analysis, there was an effect of surface frequency in block 1, $t(2699) = -2.83$, $\chi^2(1) = 8.38$, $p < .0039$, and a smaller sized effect in block 2, $t(3741) = -2.50$, $\chi^2(1) = 6.57$, $p < .0104$.

Discussion

Picture-naming latencies in Experiment 1 were sensitive to the compound's surface frequency, but not to any of the measures associated with the compound's constituents (i.e., lemma frequency, cumulative frequency, family size, positional frequency, positional entropy, complement frequency, derivational entropy). In addition, naming latencies were sensitive to the

[2] Note that frequency is manipulated between items, and hence, no by-item random slopes for frequency are possible. Furthermore, including the interaction of the frequency measures with block in the by-participant random-slopes leads to a model that is unidentified; each participant rarely contributes more than a single data point for each frequency value in a given block. These issues contributed to the failure of converge of a maximal model of random-effects structure (Barr et al., 2013). These concerns also hold for analysis 2.

[3] The reader may be surprised to observe that while latencies became shorter with increasing block, latencies became longer with increasing trial. However, this is because trial indexes the position of a trial *within* a block, and not across blocks. Hence, although latencies decreased with increasing block number, within a block, latencies increased with increasing trial number.

[4] Additional analyses using cumulative constituent frequency instead of lemma frequency yielded identical results.

Table 2 Correlation matrix reporting Pearson's correlation coefficients between the variables used in the statistical models of Experiment 1

Variables	Fam	NIA	VC	S.freq	L.lem	R.lem	L.PC	R.PC
Fam	1.00	.15	-.13	.00	-.03	.00	-.11	.11
NIA	.15	1.00	.02	.03	-.05	.06	.05	.14
VC	-.13	.02	1.00	.03	.00	.09	.10	-.01
S.freq	.16	.06	.02	1.00	.02	.00	-.05	.00
L.lem	.16	-.03	-.22	.19	1.00	-.12	-.77	-.14
R.lem	.15	.08	.06	.16	-.11	1.00	.13	.83
L.PC	-.11	.05	.10	-.26	-.79	.12	1.00	.17
R.PC	.11	.14	-.01	.23	-.11	.84	.17	1.00

Note. Below the diagonal (gray background), correlations *before* residualizing and centering; above the diagonal (white background), correlations *after* residualizing and centering. Note the absence of any correlation between the surface frequency and relevant constituent variables after residualization. Fam = familiarity, NIA = name–image agreement, VC = visual complexity, S.freq = surface frequency, L.lem = left constituent lemma frequency, R.lem = right constituent lemma frequency, L.PC = left (modifier) principal component, R.PC = right (head) principal component. Please note that the correlations between constituent lemma frequency and principal components were not residualized, given that it was not our concern to distinguish between two types of variables, and were analyzed in separate models. Note also that although a single surface frequency variable is presented here, two different residualized versions were used for the analyses of basic frequency and family size measures.

picture's familiarity, name–image agreement, and visual complexity measures (e.g., Ellis & Morrison, 1998).

A potential concern with these results is that they rely exclusively on the use of frequency measures obtained from CELEX. Recently, researchers have argued that measures obtained from other sources, such as SUBTLEX, provide more accurate estimates of word frequency than does CELEX (e.g., Brysbaert & New, 2009). Consequently, one might argue that the failure to obtain effects for the constituent frequency measures in Experiment 1 reflects the specific CELEX measures used in the experiment. To address this issue, we reanalyzed the data using surface and constituent frequency measures obtained from SUBTLEX. The results obtained in this reanalysis were identical to those obtained in Experiment 1, except that we did not find an interaction between block and surface frequency.

These results establish the reliability of the findings by Janssen et al., (2008). In picture naming, compound word production yields a surface frequency effect but no constituent effects. In addition, the frequency effect was smaller after familiarization (in block 2), which rules out the possibility that the familiarization procedure caused the surface frequency effect previously observed by Janssen et al. (2008).

As we discussed above, this pattern of results obtained in the picture-naming task is at odds with that found in other language production and comprehension tasks. For example, constituent effects have been robustly found in lexical decision tasks (e.g., Kuperman et al., 2009). Our main goal here was to rule out the possibility that the contrasting constituent effects observed between picture naming and lexical decision tasks reflect methodological differences. In Experiment 2 we examined this possibility directly by extracting lexical decision latencies for the compounds used in Experiment 1 from the English Lexicon Project (Balota et al., 2007).

Experiment 2: Lexical decision

Method

Materials

The English Lexicon Project contains a sample of over 40,000 words for which lexical decision times are available. These lexical decision times were collected in subsets from a large number of participants (>800). The participants were students at universities from various regions of the U.S. For the lexical decision task, the word stimuli were selected from Francis and Kučera and CELEX, and the pronounceable nonword stimuli were created by changing one or two letters in a corresponding target word (see Balota et al., 2007, for more details).

Lexical decision times were available for a subset of 95 words from the original set of 150 words (63 %). In this subset, the average surface frequency was higher than in the original set, $t(243) = 3.49$, $p < .001$, as was the left positional frequency, $t(243) = 3.18$, $p < .002$. No other variables in the subset differed statistically from those in the original, indicating that the subset is a representative sample of the compounds used in Experiment 1. For the lexical properties of the items in this subset, see Table 1.

Analysis and results

Given that our latencies from the English Lexicon Project were item based (i.e., averaged over participants), we conducted a standard item-based ordinary least squares regression analysis. As before, to combat collinearity, we centered and residualized variables with a Pearson correlation coefficient >.15. The correlation matrix of the variables before and after residualization is presented in Table 4. We conducted the two analyses that paralleled those of Experiment 1.

Table 3 Analysis of the picture naming latencies of Experiment 1 using mixed effect modeling

Analysis	Variable	Full Model		Model Comparisons	
		β (*SE*)	*t*(6445)	$\chi^2(1)$	*p*
Analysis 1	Intercept	1,020.05 (18.73)	54.46	468.04	.0000
	Block = 1	−239.26 (6.86)	−34.87	1,100.42	.0000
	Surface freq	−19.92 (6.77)	−2.94	8.99	.0027
	Left lemma freq	2.67 (7.23)	.37	.14	.7039
	Right lemma freq	−3.52 (7.23)	−.49	.25	.6153
	Trial	.29 (0.09)	3.23	13.43	.0002
	Familiarity	−25.65 (7.87)	−3.26	14.02	.0002
	Name-image ag	−91.27 (12.73)	−7.17	47.10	.0000
	Visual complexity	16.33 (8.17)	2.00	4.22	.0400
	Plosiveness = yes	−25.90 (15.52)	−1.67	2.95	.0861
	Fricative = yes	15.01 (17.18)	.87	.82	.3658
	Voicing = yes	2.76 (13.48)	.21	.05	.8317
	Block:surface freq	11.55 (4.08)	2.83	7.98	.0047
	Block:left lemma freq	−.79 (4.22)	−.19	.03	.8570
	Block:right lemma freq	4.38 (4.29)	1.02	1.04	.3077
Analysis 2	Intercept	1,022.21 (18.93)	54.01	463.87	.0000
	Block = 1	−239.33 (6.86)	−34.90	1,102.35	.0000
	Surface freq	−21.38 (7.06)	−3.03	9.51	.0020
	Left PC	−.63 (5.65)	−.11	.01	.9086
	Right PC	−2.08 (5.27)	−.39	.16	.6863
	Trial	.28 (0.09)	3.20	10.27	.0014
	Familiarity	−23.18 (8.08)	−2.87	8.57	.0034
	Name–image ag	−91.91 (12.75)	−7.21	47.49	.0000
	Visual complexity	16.84 (8.16)	2.06	6.07	.0138
	Plosiveness = yes	−26.99 (15.84)	−1.70	3.07	.0796
	Fricative = yes	12.74 (17.34)	.73	2.17	.1403
	Voicing = yes	1.59 (13.59)	.12	.01	.9034
	Block:surface freq	14.91 (4.21)	3.54	12.50	.0004
	Block:left PC	−2.51 (3.21)	−.78	.62	.4304
	Block:right PC	3.34 (3.19)	1.04	1.09	.2969

Note. Analysis 1 refers to the basic frequency measures, and analysis 2 refers to the family size measures. Full model refers to the results from the analysis when all predictors are simultaneously present in the model. Model comparison refers to the results from likelihood ratio tests, where a model with and without the predictor are compared. Degrees of freedom were determined by subtracting the number of fixed effect predictors from the total number of data points (Baayen, 2008). Note the effect of surface frequency and the absence of effects of constituent measures. Name–image ag = name–image agreement; freq = frequency; PC = principal component.

For the first analysis we considered the variables surface frequency and the left and right constituent lemma frequency. As can be seen in Table 5 and Fig. 2, lexical decision latencies were sensitive to the compound's surface frequency (Fig. 1a) and its left and right constituent lemma frequency (Fig. 1b, c).

In the second analysis, we included the variables surface frequency and the left and right constituent principal components associated with a compound's constituent family size. This analysis revealed effects of the surface frequency (Fig. 1d) and of the left and the right constituent principal components (Fig. 1e, f).

Discussion

The results of Experiment 2 contrast sharply with those of Experiment 1. Whereas in Experiment 1, picture-naming latencies were affected by the compound's surface frequency effect, but not its constituent measures, in Experiment 2, the lexical decision latencies were sensitive to

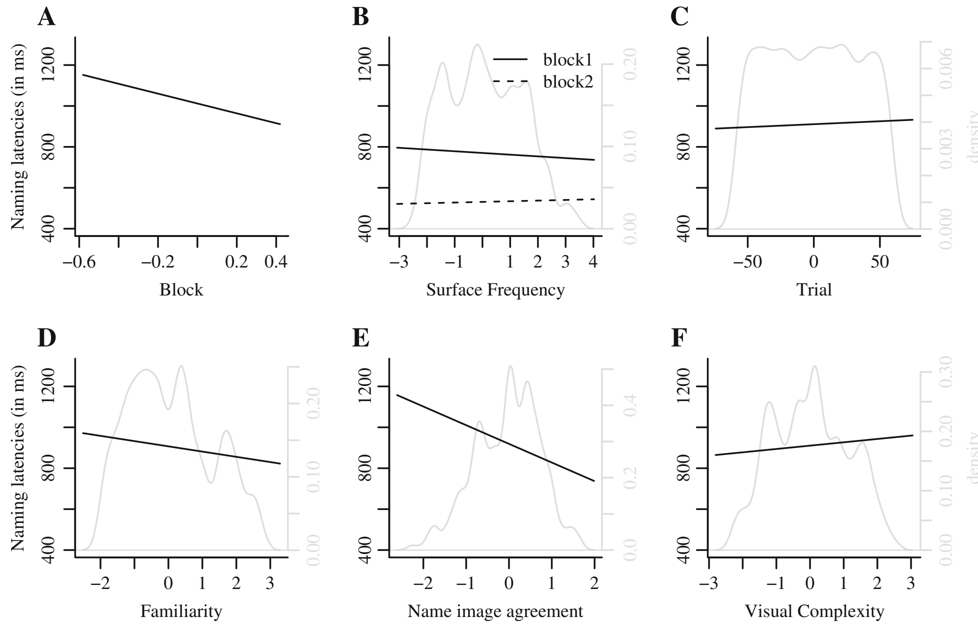

Fig. 1 Graphical presentation of the picture-naming results of Experiment 1. Each panel (a–f) shows the effect of a significant predictor on naming latencies as a function of a predictor's values on the *x*-axes. The effect of each predictor is adjusted for the effects of the other predictors. Naming latencies are shown on the left *y*-axes (in milliseconds) in black. Also plotted are the relative occurrences of each value of a continuous predictor (by means of a smoothed histogram function) on the right *y*-axes in gray

the compound's surface frequency and to its left and right constituent measures.

Two further methodological concerns need to be addressed before we can proceed to a general interpretation of the results. First, Experiment 2 contained a subset of the items used in Experiment 1. Therefore, one could argue that the contrasting results between the two experiments were caused by these different sets of items. However, a reanalysis of Experiment 1 with the items of Experiment 2 did not substantially change the pattern of results of Experiment 1. Specifically, as before, surface frequency effects were obtained in the absence of effects for the compound's constituent measures (see Table 6 for details).

Second, different types of analyses were used between Experiments 1 and 2. Experiment 1 relied on a mixed-effects technique with crossed random effects for participants and items, while Experiment 2 relied on an item-based ordinary least squares regression. To rule out that this difference in the type of analyses impacted the results, we reanalyzed the results of Experiment 1 with an ordinary least squares method. This was achieved by averaging the naming latencies across participants, thereby obtaining item-based RT averages. This least squares analysis produced the same results as those found with the mixed-effects method, except that no interaction between block and surface frequency was found (see Table 7 for details).

Table 4 Correlation matrix reporting Pearson's correlation coefficients between the variables used in the statistical models of Experiment 2 reported below.

Variables	S.freq	L.lem	R.lem	L.PC	R.PC
S.freq	1.00	-.07	.00	-.12	.00
L.lem	.08	1.00	.03	-.81	-.05
R.lem	.33	.03	1.00	.02	.82
L.PC	-.07	-.81	.02	1.00	.15
R.PC	.34	-.05	.82	.15	1.00

Note. Below the diagonal (gray background), correlations *before* residualizing and centering; above the diagonal (white background), correlations *after* residualizing and centering. Note the absence of any correlation between the surface frequency and relevant constituent variables after residualization. S.freq = surface frequency, L.lem = left constituent lemma frequency, R.lem = right constituent lemma frequency, L.PC = left (modifier) principal component, R.PC = right (head) principal component; As before, correlations between constituent lemma frequency and principal components were not residualized, and two different residualized versions of surface frequency were used for the analyses of basic frequency and family size measures.

General discussion

In Experiment 1, picture-naming latencies were sensitive to the compound's surface frequency, picture familiarity, and name–image agreement properties, but not to its constituent measures. In Experiment 2, lexical decision

Table 5 Analysis of the lexical decision latencies of Experiment 2 using ordinary least squares

Analysis	Variable	β (*SE*)	*t*(91)	*p*
Analysis 1	Intercept	718.37 (7.95)	90.30	.0001
	Surface frequency	−12.46 (5.30)	−2.35	.0208
	Left lemma frequency	−15.31 (5.76)	−2.66	.0094
	Right lemma frequency	−21.66 (4.98)	−4.35	.0001
Analysis 2	Intercept	718.37 (7.87)	91.23	.0001
	Surface frequency	−12.85 (5.28)	−2.43	.0171
	Left PC	18.20 (4.59)	3.96	.0002
	Right PC	−14.93 (3.80)	−3.93	.0002

Note. Analysis 1 refers to the analysis of the frequency variables, and analysis 2 refers to the analysis of the family size variables. Note the presence of both surface and constituent effects. PC = principal component.

latencies were sensitive to the compound's surface frequency, left and right constituent frequency, and family size measures.

These results confirm the contrasting effects of constituent measures in the picture-naming and lexical decision tasks previously found in the literature. For example, Janssen et al. (2008) reported surface frequency but not constituent effects in a picture-naming task with compound words, whereas Juhasz et al. (2003; Kuperman et al., 2009) reported surface frequency and constituent effects in a lexical decision task. These contrasting constituent effects were also obtained in our picture-naming and lexical decision tasks, where the same methodologies (materials, language, analyses) were used. Thus, methodological differences alone cannot account for the observed contrast.

Instead, we propose that this pattern of contrasting constituent effects arises due to differences in the degree to which the input representation in a task transparently contains the compound's constituents. Consider first word comprehension tasks such as lexical decision. In this task, the input representation is an orthographic or phonological string. This input string transparently contains the compound word's constituents, meaning that they are directly recoverable from the physical signal. For such input representations, it is assumed that they lead to the activation of the compound's constituents in the lexicon, which in turn leads to the observed constituent effects. By contrast, in the picture-naming task, a semantic input representation drives word retrieval (e.g., Potter & Faulconer, 1975). Unlike the orthographic or phonological input representations in word comprehension tasks, a semantic input representation does not transparently contain a compound's constituents. For such input representations, it is assumed that they do not directly activate the constituents in the lexicon, leading to the absence of constituent effects.[5] Thus, the presence of constituent effects in a task depends on the degree to which the input representation in a task transparently contains the compound's constituents.

The absence of constituent effects in the picture-naming task reported here contrasts with the presence of such effects in other word production tasks, such as the response association and picture–word interference tasks (Bien et al., 2005; Roelfs, 1996; Zwitserlood et al., 2000). Although our study was not designed to resolve this issue, we discuss two possible reasons for this inconsistency. First, the observed discrepancy could be the result of the different languages used: Whereas constituent effects have been reported in studies using German and Dutch speakers, no constituent effects have been reported in studies using English and Chinese speakers. Cross-linguistic differences may influence compound representation and account for the contrasting results in the reported studies. However, note that this explanation assumes that compound representation is more similar between English and Chinese than between English and Dutch, which seems unlikely (see Janssen et al., 2008, for a similar argument). A more likely possibility is that compound processing differs between tasks, where some tasks emphasize the processing of the compound's constituents, whereas other tasks do not. Consider, for example, the observation that compound word production yields robust surface frequency effects in the picture-naming task, but not in the response association task (Bien et al., 2005, exp4). This contrast in the presence of the surface frequency effect clearly points to underlying differences between the two tasks that may affect the processing of compound words. Future studies should be geared towards resolving this issue.[6]

Other studies have also used the picture-naming task to investigate the production of complex words other than compounds. These studies have reported both surface and constituent effects in the production of singular and plural inflected nouns (Baayen, Levelt, Schreuder, & Ernestus, 2007) and regular and irregular verbs (Tabak, Schreuder, & Baayen, 2010). These data are fully compatible with the proposal advanced here, if one further nuance is taken into account. This concerns the assumption that the nature of the input representation in a task is determined not only by the task, but by a combination of the task and the specific type of the

[5] Compound constituents may be activated weakly because of shared semantic representations, just as would be the case for any two semantically related lexical entries.

[6] Constituent effects have also been observed in various versions of the picture–word interference task (Zwitserlood et al., 2000). However, as was discussed by Janssen et al. (2008), these effects may reflect semantic and not morphological priming. Specifically, the interpretation of long-lag morphological priming effects rests on the absence of long-lag semantic priming effects (Feldman, 2000). However, long-lag semantic priming effects are reported in the literature (e.g., Becker, Moscovitch, Behrmann, & Joordens, 1997), thereby undermining a nonsemantic interpretation of long-lag morphological priming effects. See Janssen et al. (2008) for further discussion of this issue.

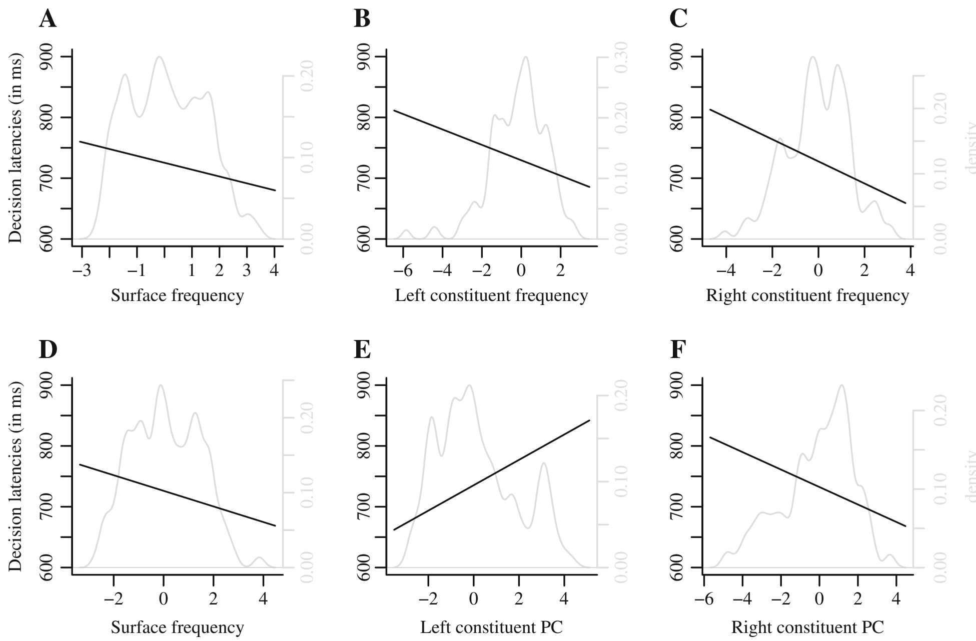

Fig. 2 Graphical presentation of the lexical decision results of Experiment 2, which considered the effects of surface and left and right constituent frequency (top row) and the effects of surface frequency and left and right constituent PCs (bottom row) in two separate analyses. The effect of each predictor is adjusted for the effects of the other predictors. Decision latencies are shown on the left *y*-axes (in milliseconds) in black, and the relative occurrences of each value of a continuous predictor are shown on the right y-axes in gray

complex word that is used in the experiment. In other words, it is not our proposal that the task per se determines whether the constituents are activated in the lexicon. Rather, our proposal is that both the type of complex word and the task together determine the nature of the input representation and that this, in turn, determines the activation of the constituents. With respect to the constituent effects found in the picture-naming studies with inflected nouns and verbs, we make the plausible assumption that the semantic representation that underlies plural inflected nouns and verbs transparently contains the complex word's constituents, leading to the observed constituent effects. This would then lead to the testable prediction that in a picture-naming task with compound word production, constituent effects should be modulated by the semantic transparency of the compound (Libben et al., 2003).

Our results challenge the notion that morphology plays a central and obligatory role in the retrieval of words from the lexicon. Models that incorporate this idea have been proposed in both the word comprehension (e.g., Taft, 2003; Taft & Ardasinski, 2006), and word production (e.g., Levelt, Roelofs, & Meyer, 1999) literature. These models assume that a word's lexical representation involves a decomposed morpheme-level representation and that access to these morpheme-level representations is a *necessary* part of word retrieval. The contrasting constituent effects in compound word processing between picture-naming and lexical decision tasks observed here present problems for models that rely on these assumptions. Specifically, if access to a compound's constituents were obligatory for the retrieval of the compound word from the lexicon, one would expect effects of the compound's constituents in all tasks. The observation that constituent effects were observed in lexical decision, but not in picture naming, therefore undermines the notion that morphology plays a central and obligatory role in lexical access.

In the word comprehension literature, alternative models that reject the idea of obligatory access have been proposed. Specifically, such models assume that the recognition of morphologically complex words relies on two parallel routes: a whole-word route in which the access to meaning takes place on the basis of the whole-word form and a decomposed route in which the meaning of the complex word is accessed on the basis of its constituents (e.g., Caramazza et al., 1988; Schreuder & Baayen, 1997). A dual-route-type architecture could account for the results observed here by assuming that the degree to which the decomposed route is activated depends on the degree to which the input representation of a task transparently contains the compound's constituents. In the lexical decision task, the decomposed route is strongly activated on the basis of the transparent input representation, leading to the observed constituent effects. In a picture-naming task with compound word stimuli, the decomposed route is weakly or not activated, leading to the absence of constituent effects.[7]

One additional aspect of our data deserves further discussion. In the picture-naming task of Experiment 1, we did not observe any effects of the measures associated with the compound's left and right constituent family size. This is surprising, considering that family size effects have been interpreted

Table 6 Model statistics of the variables in a reanalysis of Experiment 1 with the 95 items used in Experiment 2

		Full Model		Model Comparisons	
Analysis	Variable	β (*SE*)	*t*(4236)	$\chi^2(1)$	*p*
Analysis 1	Intercept	966.02 (25.32)	38.14	275.67	.0000
	Block = 1	−232.20 (8.34)	−27.81	701.38	.0000
	Surface freq	−15.58 (9.69)	−1.61	4.01	.0452
	Left lemma freq	−2.82 (11.18)	−.25	.07	.7912
	Right lemma freq	−3.02 (9.11)	−.33	.12	.7279
	Trial	.37 (0.11)	3.51	12.11	.0005
	Familiarity	−26.76 (10.11)	−2.64	7.52	.0061
	Name–image ag	−92.14 (15.66)	−5.88	33.72	.0000
	Visual complexity	13.22 (10.44)	1.27	1.77	.1829
	Plosiveness = yes	1.08 (20.03)	.05	.00	.9544
	Fricative = yes	48.64 (23.55)	2.06	5.83	.0157
	Voicing = yes	12.32 (18.02)	.68	.67	.4135
	Block:surface freq	9.97 (5.42)	1.84	3.41	.0648
	Block:left lemma freq	−.49 (6.17)	−.08	.01	.9301
	Block:right lemma freq	6.69 (5.18)	1.29	2.76	.0969
Analysis 2	Intercept	964.07 (25.15)	38.33	276.78	.0000
	Block = 1	−232.00 (8.34)	−27.80	700.87	.0000
	Surface freq	−18.18 (9.68)	−1.88	3.30	.0692
	Left PC	−5.21 (8.97)	−.58	.81	.3668
	Right PC	2.63 (6.86)	.38	.17	.6787
	Trial	.37 (0.11)	3.45	11.13	.0008
	Familiarity	−28.48 (10.20)	−2.79	7.77	.0053
	Name–image ag	−92.58 (15.53)	−5.96	32.69	.0000
	Visual complexity	12.96 (10.31)	1.26	1.74	.1874
	Plosiveness = yes	−.37 (19.63)	−.02	.00	.9999
	Fricative = yes	52.45 (23.72)	2.21	4.74	.0294
	Voicing = yes	12.82 (17.78)	.72	1.03	.3108
	Block:surface freq	12.59 (5.45)	2.31	5.73	.0167
	Block:left PC	.93 (4.87)	.19	.00	.9999
	Block:right PC	−1.46 (4.01)	−.36	.00	.9999

Note. Note the absence of any constituent effects. Name–image ag = name–image agreement; freq = frequency; PC = principal component.

not at a morphological, but at a semantic level (e.g., Dijkstra, Moscoso del Prado Martin, Schulpen, Schreuder, & Baayen, 2005; Moscoso del Prado Martin, Bertram, Häikiö, Schreuder, & Baayen, 2004). Given that the picture-naming task is sensitive to the processing of semantic information, one would have expected family size effects in this task. One possible explanation of this state of affairs is within the dual-route view proposed above. One might assume that the family size effect does not reflect any intrinsic aspect of the meaning of a complex word but that its effects arise from the decomposed processing of the complex word's constituents. Under such an assumption, one would expect no effect of a compound's constituent family size in a picture-naming task, but one would expect such effects in other tasks in which the input representation allows for decomposed retrieval. Such a proposal seems borne out by the data, where family size effects have been found in response association (Bien et al., 2005) and comprehension tasks such as lexical decision and naming (Juhasz & Berkowitz, 2011). Further targeted studies should confirm this interpretation.

Finally, the present results add to a growing body of data that show a failure to detect constituent effects in compound

[7] Connectionist models (Baayen, Milin, Durdević, Hendrix, & Marelli, 2011; Rueckl, & Raveh, 1999; Seidenberg & Gonnerman, 2000) also reject the assumption of obligatory access and, therefore, could, in principle, also account for the results presented here.

Table 7 Model statistics of the variables in a reanalysis of Experiment 1 using the ordinary least squares regression method used in Experiment 2

		Full Model		
Analysis	Variable	β (*SE*)	*t*(285)	*p*
Analysis 1	Intercept	1,059.95 (16.82)	63.00	.0001
	Block = 1	−264.81 (19.82)	−13.36	.0001
	Surface freq	−20.86 (6.02)	−3.47	.0007
	Left lemma freq	4.18 (6.26)	0.67	.5052
	Right lemma freq	−5.99 (6.40)	−0.94	.3494
	Familiarity	−29.72 (7.02)	−4.23	.0001
	Name–image ag	−99.45 (11.46)	−8.67	.0001
	Visual complexity	17.92 (7.28)	2.46	.0145
	Plosiveness = yes	−22.65 (13.88)	−1.63	.1039
	Fricative = yes	19.78 (15.24)	1.30	.1953
	Voicing = yes	−0.49 (12.00)	−0.04	.9669
	Block:surface freq	11.03 (12.03)	.92	.3602
	Block:left lemma freq	−3.34 (12.35)	−0.27	.7868
	Block:right lemma freq	7.53 (12.69)	.59	.5534
Analysis 2	Intercept	1,062.17 (17.01)	62.44	.0000
	Block = 1	−264.82 (19.79)	−13.38	.0001
	Surface freq	−22.93 (6.29)	−3.65	.0003
	Left PC	−2.38 (4.83)	−0.49	.6224
	Right PC	−2.56 (4.67)	−0.55	.5842
	Familiarity	−27.19 (7.19)	−3.78	.0002
	Name–image ag	−100.15 (11.47)	−8.73	.0001
	Visual complexity	18.23 (7.27)	2.51	.0127
	Plosiveness = yes	−24.14 (14.13)	−1.71	.0888
	Fricative = yes	17.23 (15.38)	1.12	.2635
	Voicing = yes	−2.11 (12.09)	−0.18	.8615
	Block:surface freq	16.68 (12.47)	1.33	.1823
	Block:left PC	.86 (9.30)	.09	.9261
	Block:right PC	4.88 (9.13)	.54	.5934

Note. Note that this analysis does not require model comparisons to obtain *p*-values and that the variable trial could not be included. Note also the persistent surface frequency effect in the absence of a constituent effect. Name–image ag = name–image agreement; freq = frequency; PC = principal component.

word production. Besides the data of Janssen et al. (2008), two recent studies have also failed to find effects of constituents in compound word production. First, Cohen-Goldberg (2013) examined the duration of compound word latencies in natural speech selected from the Buckeye corpus (Pitt et al., 2007). Using regression analyses, Cohen-Goldberg found that compound word duration latencies were predicted by the compound's surface frequency, but not by its constituent frequency. Likewise, Jacobs and Dell (in press) failed to obtain a second constituent priming effect during compound word production. Specifically, phonological priming effects patterned similarly for compounds and bisyllabic monomorphemic words and differed from those found for two-word utterances. Both Cohen-Goldberg and Jacobs and Dell have interpreted these data in terms of a model that assumes whole-word representations for compounds at the phonological level. These independently collected data points instill further confidence that the failure to detect the constituent effects in the presenst study is a real null effect.

To conclude, the primary motivation of our study was to resolve a surprising discrepancy in the literature: The observation that whereas the majority of tasks reveal constituent effects in compound processing, the picture-naming task does not. Our results establish that these contrasting constituent effects cannot be explained by methodological differences alone. Instead, we propose that these contrasts are more likely the result of task-specific differences in the degree to which the input representation in a given task transparently contains the compound's constituents. Our results are not readily accommodated by models that assume that access to a compound's constituents is an obligatory aspect of compound word retrieval (Levelt et al., 1999; Taft, 2003) but find a more straightforward explanation in models that assume parallel dual routes to constituent access (Caramazza et al., 1988; Schreuder & Baayen, 1997), where the activation of the decomposed route depends on the nature of the input representation in the task. An avenue for future research would be to examine the role of semantic transparency in the production of compounds in the picture-naming task.

Acknowledgments This research was supported by Grant RYC-2011-08433 from the Spanish Ministry of Economy and Competitiveness to N.J. and by Grant R01DC006842 from the National Institute on Deafness and Other Communication Disorders and the Societã Mente/Cervello, Fondazione Cassa di Risparmio di Trento e Rovereto to A.C.

References

Andrews, S. (1986). Morphological influences on lexical access: Lexical or nonlexical effects? *Journal of Memory and Language, 25*(6), 726–740.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge Univ Pr.

Baayen, R. (2010). A real experiment is a factorial experiment. *The Mental Lexicon, 5*(1), 149–157.

Baayen, R., Piepenbrock, R., & van Rijn, H. (1993). The *{CELEX}* lexical data base on *{CD – ROM}*.

Baayen, R., Levelt, W., Schreuder, R., & Ernestus, M. (2007). Paradigmatic structure in speech production. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (pp. 1–29, vol 43). Chic Ling Society.

Baayen, R., Milin, P., Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438–481.

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316.

Balota, D., Yap, M., Hutchison, K., Cortese, M., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods,* 39(3), 445–459.

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

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1059–1082.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long finnish compounds. *Journal of Memory and Language, 48*(3), 615–634.

Bien, H., Levelt, W., & Baayen, R. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the United States of America, 102*(49), 17876–17881.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods, 41*(4), 977–990.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*(3), 297–332.

Cohen-Goldberg, A. M. (2013). *The measure of a (post)man: Investigating the lexical representation of compounds in production. Talk presented at the XIth International Symposium of Psycholinguistics*. Spain: Tenerife.

De Jong, N., Feldman, L., Schreuder, R., Pastizzo, M., & Baayen, R. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language, 81*(1–3), 555–567.

Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology and skilled reading: An empirical overview.

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

Dohmes, P., Zwitserlood, P., & Bölte, J. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language, 90*(1–3), 203–212.

Duñabeitia, J., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? evidence for morpho-orthographic decomposition. *Cognition, 105*(3), 691–703.

Ellis, A., & Morrison, C. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 515–523.

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.

Forster, K., & Forster, J. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods, 35*(1), 116–124.

Glaser, W. (1992). Picture naming. *Cognition, 42*(1–3), 61–105.

Gumnior, H., Bölte, J., & Zwitserlood, P. (2006). A chatterbox is a box: Morphology in german word production. *Language and Cognitive Processes, 21*(7–8), 920–944.

Jacobs, C. L., & Dell, G. S. (in press). “hotdog”, not “hot dog”: The phonological planning of compound words. Language, Cognition, and Neuroscience

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for mandarin chinese and english compounds. *Language and Cognitive Processes, 23*(7–8), 1191–1223.

Janssen, N., Pajtas, P., & Caramazza, A. (2011). A set of 150 pictures with morphologically complex english compound names: Norms for name agreement, familiarity, image agreement, and visual complexity. *Behavior Research Methods, 43,* 478–490.

Juhasz, B. (2008). The processing of compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes, 23*(7–8), 1057–1088.

Juhasz, B., & Berkowitz, R. (2011). Effects of morphological families on english compound word recognition: A multitask investigation. *Language and Cognitive Processes, 26*(4–6), 653–682.

Juhasz, B., Starr, M., Inhoff, A., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology, 94*(2), 223–244.

Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *NeuroImage, 42*(4), 1622–1630.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. (2009). Reading polymorphemic dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 876–895.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

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

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. (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), 1271–1278.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University.

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*(2), 820–833.

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

Roelofs, A. (1996). Morpheme frequency in speech production: Testing weaver. *Yearbook of Morphology, 1996,* 135–154.

Rueckl, J., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language, 68*(1–2), 110–117.

Schreuder, R., & Baayen, R. (1997). How complex simplex words can be. *Journal of Memory and Language, 37,* 118–139.

Seidenberg, M., & Gonnerman, L. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*(9), 353–361.

Shannon, C. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3–55.

Snodgrass, J., & 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*(2), 174–215.

Tabak, W., Schreuder, R., & Baayen, R. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon, 5*(1), 22–46.

Taft, M. (2003). Morphological representation as a correlation between form and meaning. *Reading Complex Words, 1,* 113–137.

Taft, M., & Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *The Mental Lexicon, 1*(2), 183–199.

Taft, M., & Forster, K. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 607–620.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production: Evidence from picture naming. *Language and Cognitive Processes, 15*(4–5), 563–591.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2002). Where and how morphologically complex words interplay with naming pictures. *Brain and Language, 81*(1–3), 358–367.