Specific-Word Frequency Is Not All That Counts in Speech Production: Comments on Caramazza, Costa, et al. (2001) and New Experimental Data

Jörg D. Jescheniak
Max Planck Institute of Cognitive Neuroscience
and University of Potsdam

Antje S. Meyer
University of Birmingham

Willem J. M. Levelt
Max Planck Institute for Psycholinguistics

A. Caramazza, A. Costa, M. Miozzo, and Y. Bi (2001) reported a series of experiments demonstrating that the ease of producing a word depends only on the frequency of that specific word but not on the frequency of a homophone twin. A. Caramazza, A. Costa, et al. concluded that homophones have separate word form representations and that the absence of frequency-inheritance effects for homophones undermines an important argument in support of 2-stage models of lexical access, which assume that syntactic (lemma) representations mediate between conceptual and phonological representations. The authors of this article evaluate the empirical basis of this conclusion, report 2 experiments demonstrating a frequency-inheritance effect, and discuss other recent evidence. It is concluded that homophones share a common word form and that the distinction between lemmas and word forms should be upheld.

According to current two-stage models of lexical access in speech production (e.g., Dell, 1990; Dell & O'Seaghdha, 1992; Garrett, 1988; Levelt, Roelofs, & Meyer, 1999; Levelt et al., 1991; see Levelt, 1999, for a review), the representations of words in the mental lexicon come in two kinds. On the one hand, there are syntactic representations (referred to as *lemmas*), and on the other hand, there are phonological, or word form, representations (often referred to as *lexemes*, see Kempen & Huijbers, 1983). Evidence that is consistent with this view comes from such divergent sources as the distributional properties of spontaneously occurring and laboratory-induced speech errors, chronometric behavioral data, and electrophysiological measures. However, Caramazza and colleagues have repeatedly pointed out that some of these data can also be explained without invoking a separate lemma level (e.g., see Caramazza, 1997; Caramazza & Miozzo, 1998; Roelofs, Meyer, & Levelt, 1998, for further discussion). Here, we do not summarize all of these findings but focus on one observation, namely the frequency-inheritance effect in the production of homophones (Dell, 1990; Jescheniak & Levelt, 1994).

The term *frequency inheritance* refers to the observation that the ease of producing a word is affected by the existence of another word with the same phonological form. More specifically, a low-frequency word (e.g., *butt*, occurring only rarely in the language) with a high-frequency homophone twin (e.g., *but*, occurring very often) is produced about as easily as a control word matched in frequency to the sum frequency of the low-frequency word and its twin. The effect was originally reported by Dell (1990), who observed that an item's susceptibility to phonological errors was determined by the summed frequency of the item and its high-frequency homophone, rather than only by its own (specific) frequency. Jescheniak and Levelt (1994) extended the finding to the chronometric domain by demonstrating that the same pattern was obtained when the speed of producing the word in response to a second-language cue word was measured. Dell as well as Jescheniak and Levelt took these findings to support the notion of partially overlapping lexical representations for homophonic words. In particular, the findings suggest that the low-frequency target and its high-frequency twin, although having distinct lemma representations (in our example, *butt−but*, one is a noun, and the other is a conjunction), project onto a single phonological representation.

In a number of articles, Caramazza and colleagues (e.g., Caramazza, 1997; Caramazza & Miozzo, 1998) have argued against the distinction between lemmas and word form representations. In a recent article, Caramazza, Costa, Miozzo, and Bi (2001) questioned the need for postulating lemma representations on the basis of their failure to replicate the frequency-inheritance effect in homophones. In this article, we first discuss the evidence presented by Caramazza, Costa, et al. and then report two new experiments investigating frequency-inheritance effects in Dutch and German.

In Experiment 1A, Caramazza, Costa, et al. (2001) asked English-speaking participants to perform a picture-naming task.

The design was similar to the one used by Jescheniak and Levelt (1994). There were three sets of pictures: (a) pictures with low-frequency names without homophones (LF controls), (b) pictures with high-frequency names without homophones (HF controls), and (c) pictures with low-frequency names and high-frequency homophones (homophones). For the last-mentioned items, the specific frequency count was matched to the low-frequency control items, and the summed frequency count was matched to the high-frequency control items. Caramazza, Costa, et al. (2001) found that HF controls were named faster than LF controls and homophones with no significant difference between the latter two sets. This pattern is clearly at odds with what one would have predicted on the basis of Dell's (1990) and Jescheniak and Levelt's findings. However, there are a number of points to consider. First, the frequency contrast in the experiment carried out by Caramazza, Costa, et al. (factor of 4.5) was much smaller than in Jescheniak and Levelt's experiment (factor of 37.5). Consequently, LF controls and HF controls yielded a relatively small frequency effect (38 ms) in the naming experiment compared with the effect observed by Jescheniak and Levelt (99 ms). In addition, there was a 10-ms difference between high-frequency and low-frequency items in Experiment 1C, which was a delayed naming experiment determining potential articulatory contributions to the effect. The authors did not report whether the residual 28-ms effect was reliable. If it was not, then the basic frequency manipulation was not effective, and the issue of frequency inheritance could not be addressed. Second, the interpretation of Experiment 1B, determining potential contributions from object recognition processes, is complicated by a number of methodological issues. In this experiment, picture-naming latencies in Italian were assessed, the motivation being that the pictures that had homophonic names in English did not have such names in Italian. The problems here are that (a) these naming latencies provide no direct measure of the ease of conceptual processing in Experiment 1A as they were also determined by lexical variables specific to Italian, (b) the experiment did not use the same item set as Experiment 1A (15 out of 75 items were discarded), and (c) the experiment was performed on a participant sample drawn from a population for which object familiarity parameters may have been different from those characterizing the sample tested in Experiment 1A. This possibility is not all that unlikely, given that the authors excluded some items used in Experiment 1A because the authors considered them to be unfamiliar to Italian participants.

In Experiment 2A, speakers of Mandarin Chinese performed a picture-naming task. All experimental items were homophones, and in each case the low-frequency twin was depicted. The three item sets differed in the specific frequency of the depicted objects (46, 61, and 737 occurrences in a corpus of unspecified size, for Sets 1 to 3) and in the frequency ratio of the nondepicted high-frequency twin to the depicted low-frequency element (ratios of 28.9, 1.9, and 2.6). Naming latencies amounted to 783, 749, and 717 ms for Sets 1, 2, and 3, respectively. Surprisingly, the items of Set 1, which had high-frequency homophone twins, were named more slowly than the items of Set 2, which did not have high-frequency twins. As a control for contributions from object recognition, Caramazza, Costa, et al. again asked speakers of a different language (English in this case) to name the pictures. In English, the picture names were not homophones, and the items of Sets 1 and 2 did not differ in name frequency, but the frequency of

the items in Set 3 was higher, as in the Chinese experiment. The authors obtained exactly the same pattern of results as in the Chinese experiments ($M = 737$, 702, and 651 ms for Sets 1 through 3). Caramazza, Costa, et al. proposed that the difference between Set 1 and Set 2, found both in the Chinese and in the English experiment, was because of a difference in the ease of recognizing the pictures, whereas the difference between Set 2 and Set 3 was related to a difference in the frequency of the object names. However, the data do not rule out the possibility that all latency differences obtained in these experiments arose because the depicted objects differed in ease of object recognition. In any event, there was no evidence for homophone frequency inheritance.

Experiment 3A is most comparable with the experiment carried out by Jescheniak and Levelt (1994). It tested English–Spanish bilinguals, using a translation task with Spanish (second language) probe words and English (native language) target words. The frequency contrast between the high-frequency and the low-frequency control items was even larger than in the original experiment by Jescheniak and Levelt (factor of 78.0). This time there was a large frequency effect in translation latencies (206-ms difference). It is important to note that the mean latency for the homophones, which were matched with the low-frequency control items for specific frequency and with the high-frequency control items for cumulative frequency, was very similar to the latency for the low-frequency control items. Homophones also behaved like LF controls in the lexical-decision task used in Experiment 3B, which was introduced to control for contributions of probe word recognition to overall naming latencies. However, the results of this experiment are difficult to evaluate because a sample of participants from a different population (Spanish monolinguals) was tested.

Although we do not know the exact impact of these methodological and empirical issues, they complicate the interpretation of the data presented by Caramazza, Costa, et al. (2001). Still, the data provide a challenge to our view of the lexical representation of homophones. This holds in particular for Experiment 3, as it used the same experimental procedure with which we had observed a sizable frequency-inheritance effect for homophones in Dutch. In view of the conflicting findings, we decided to assess the reliability of our earlier observation by performing two experiments involving the translation paradigm. These experiments are described next. We first report an exact replication of Experiment 6 reported by Jescheniak and Levelt (1994) with Dutch as the target language, and we then report a similar experiment using German as the target language. As in our earlier experiment (Jesheniak & Levelt, 1994, Experiment 6), translation latencies and semantic decision latencies (controlling for possible contributions from differences in processing of the probes across word sets) were obtained from the same participants in two separate sessions.

Experiment 1

In Experiment 1, we report an exact replication of Experiment 6 reported by Jescheniak and Levelt (1994) with Dutch as the target language. In Experiment 2, we report a similar experiment using German as the target language.

Method

Participants. Sixteen native speakers of Dutch were recruited from the participant pool of the Max Planck Institute for Psycholinguistics. They were paid dfl 25 (approximately U.S. $10) for participating in the two experimental sessions. In this and the next experiment, all participants had normal or corrected-to-normal vision and at least 6 years of experience in reading and speaking English.

Materials. The materials were identical to those used by Jescheniak and Levelt (1994, Experiment 6). There were three sets of experimental words consisting of 11 words each, all denoting inanimate entities: (a) homophones, (b) LF controls, and (c) HF controls. The homophones were matched in lemma frequency with the LF controls, and the HF controls were matched in frequency to the sum frequency of the (low-frequency) homophones and their high-frequency twins. All three sets were matched with respect to word length in terms of phonological segments. In addition to the experimental items, a set of 33 filler items denoting animate entities were used. A set of 20 additional items, half of them denoting animate entities and half of them denoting inanimate entities, were used in a practice block. Each item was presented three times. For further details on the materials and the criteria for the preparation of the experimental lists, see Jescheniak and Levelt (1994).

Design. The design was identical to that used by Jescheniak and Levelt (1994). There were two completely crossed variables: condition (homophones vs. LF controls vs. HF controls) and repetition (1 to 3).

Procedure. The experiment consisted of two parts. Each participant was tested individually in two sessions lasting about 1 hr each. The two sessions were separated by an average of 1 week. In the first session, participants carried out the translation task. In the second session, they performed the semantic decision task by pressing one of two push buttons to indicate whether the picture showed an animate or inanimate entity. All procedural details were identical to those in the original experiment by Jescheniak and Levelt (1994).

Results and Discussion

Table 1 displays the mean translation latencies, mean semantic decision latencies, and error rates for each item set and repetition.

Table 1

Mean Translation and Semantic Decision Latencies (in ms) and the Respective Error Rates (in Percentages) by Condition and Repetition for Experiment 1

	Condition					
	Homophones		LF controls		HF controls	
Repetition	M	%	M	%	M	%
	Translation					
1	837	5.1	927	7.4	777	6.3
2	755	3.4	845	2.3	711	3.4
3	744	1.1	805	4.0	700	2.8
Average	778	3.2	858	4.6	729	4.2
	Semantic decision					
1	631	0.6	628	4.6	615	1.7
2	615	2.3	591	2.8	574	4.0
3	596	1.7	571	2.3	566	0.6
Average	614	1.5	596	3.2	585	2.0

Note. LF = low frequency; HF = high frequency.

These data are listed for the purpose of illustration only; the main statistical analyses were based on difference scores. These were obtained by subtracting each participant's semantic decision latency from that participant's translation latency for each item at each level of the repetition variable (see Jescheniak & Levelt, 1994, for details). Of the difference scores, 6.1% were missing because the categorization or translation response or both were incorrect or had occurred with a latency exceeding 2,000 ms. The averaged difference scores were submitted to analyses of variance using the fixed-variables condition (homophone, LF control, HF control) and repetition (repetition 1 to 3). Two complementary analyses were computed, one treating participants and one treating items as random variables (Clark, 1973).

The left panel of Figure 1 reproduces the mean difference scores reported by Jescheniak and Levelt (1994, Experiment 6). The right panel of Figure 1 shows the difference scores obtained in the present experiment. This data pattern exhibits three features. First, there was a significant effect of condition, $F_1(2, 30) = 73.30$, $MSE = 2,607$, $p < .01$; $F_2(2, 30) = 8.07$, $MSE = 15,234$, $p < .01$. The mean difference score was largest for LF controls (262 ms), intermediate for homophones (164 ms), and smallest for HF controls (144 ms). Second, the difference scores decreased over repetitions, $F_1(2, 30) = 12.47$, $MSE = 2,982$, $p < .01$; $F_2(2, 60) = 23.24$, $MSE = 1,217$, $p < .01$. Third, the condition effect was unaffected by repetition, $F_1(4, 60) = 1.19$, $MSE = 2,186$, $p = .33$; $F_2(4, 60) = 1.14$, $MSE = 1,217$, $p = .35$. Post hoc analyses of the condition effect (Newman–Keuls test with $p < .05$) revealed that homophones and HF controls differed from LF controls. It is most important to note that there was no significant difference between homophones and HF controls.

The error rates decreased over repetitions, $F_1(2, 30) = 4.35$, $MSE = 0.00$, $p < .05$; $F_2(2, 30) = 3.20$, $MSE = 0.00$, $p = .05$. No other effect was significant in the analyses of error rates.

In summary, low-frequency homophones were produced faster than LF controls and as quickly as HF controls, providing a full replication of Experiment 6 reported in Jescheniak and Levelt (1994). Experiment 2 tested whether this pattern generalized to a different set of materials and a different target language. In this experiment, speakers of German translated English probe words into their native language.

Experiment 2

Method

Participants. Twenty-two native speakers of German, most of them students of the University of Leipzig, were tested. For participation in the two experimental sessions, they were paid DM 26 (approximately U.S. $12). The data from 3 additional participants with excessive error rates in the translation task were discarded from the analyses.

Materials. There were again three sets of experimental words consisting of nine words each: (a) homophones, (b) LF controls, and (c) HF controls. All experimental words were monosyllabic nouns denoting inanimate entities. Items with substantial form overlap between the English probes and the German targets were not included. Each (low-frequency) homophone and its high-frequency twin differed in lemma frequency by at least a factor of 10.0. The homophones were matched in lemma frequency with the LF controls: 18.7 (*SD* = 20.5) versus 17.2 (*SD* = 24.8) mean occurrences per million words, $t(16) = 0.13$, $p = .90$. The HF controls were of similar frequency as the sum frequency of the (low-frequency) homophones and their high-frequency twins. Although a perfect match was
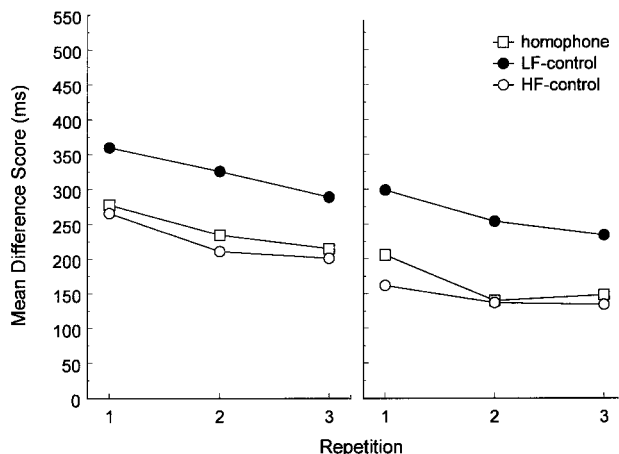
Figure 1. Mean difference scores (in ms) by condition and repetition from Jescheniak and Levelt (1994, Experiment 6; left panel) and Experiment 1 (right panel). LF = low frequency; HF = high frequency.

not possible, there was no statistical difference between the two sets: 757.6 (SD = 827.7) versus 692.1 (SD = 614.4) mean occurrences per million words, $t(16) = 0.19$, $p = .85$. The three sets of target words were matched with respect to word length in terms of phonological segments: 3.22 (SD = 0.83) for the homophones and 3.11 (SD = 0.33 for both sets) for the other two sets. In addition to the experimental items, a set of 27 filler items denoting animate entities were selected. The appendix lists all target–probe pairs.

The 27 experimental items and 27 filler items were assigned to three blocks with 18 items each, such that each block contained three items from each experimental condition and nine filler items. Two additional items (one animate and one inanimate) were selected to be used in warm-up trials at the beginning of each block. Another 18 items, half denoting inanimate entities and half denoting animate entities, were selected to be used in a practice block. Within each block, each item was presented three times. The trials were pseudorandomized with the same constraints that were applied in Experiment 1. Three different randomizations were created for each experimental block. Each participant received one version of each experimental block, preceded by the practice block. The three experimental blocks were presented in six different orders. Four of these orders were presented to 4 participants each, and two were presented to 3 participants each. Each participant received the same order of blocks in both parts of the experiment.

Design and procedure. The design and procedure were identical to those used in Experiment 1. The only exception was that probe words were presented in black Courier typeface on a light-gray background (rather than in white Arial typeface on a black background).

Results and Discussion

Table 2 displays mean translation latencies, mean semantic decision latencies, and error rates broken down by item set and repetition. As in Experiment 1, the main statistical analyses were based on difference scores, which are presented in Figure 2. Of the difference scores, 8.5% were missing because of missing translation and/or semantic decision scores.

The data pattern exhibits three features. First, there was a significant effect of condition, $F_1(2, 42) = 126.75$, $MSE = 5,919$, $p < .001$; $F_2(2, 24) = 11.11$, $MSE = 26,715$, $p < .001$. The mean difference score for LF controls was largest (380 ms), and the

Table 2

Mean Translation and Semantic Decision Latencies (in ms) and the Respective Error Rates (in Percentages) by Condition and Repetition for Experiment 2

| | Condition | | | | | |
| | Homophones | | LF controls | | HF controls | |
Repetition	M	%	M	%	M	%
			Translation			
1	930	13.6	1,125	17.2	801	4.6
2	830	5.1	962	6.6	711	1.5
3	796	3.5	923	5.6	703	1.0
Average	852	7.4	1,003	9.8	738	2.4
			Semantic decision			
1	652	2.5	688	5.1	612	3.0
2	583	1.5	598	2.0	551	.5
3	563	2.0	585	1.0	548	.5
Average	599	2.0	624	2.7	570	1.3

Note. LF = low frequency; HF = high frequency.

mean difference score for HF controls was smallest (168 ms), whereas the mean difference score for homophones was intermediate (251 ms). Second, difference scores decreased over repetitions, $F_1(2, 42) = 11.19$, $MSE = 5,746$, $p < .001$; $F_2(2, 48) = 13.93$, $MSE = 2,005$, $p < .001$. Third, the condition effect was unaffected by repetition, $F_1(4, 84) = 1.90$, $MSE = 3,647$, $p = .12$; $F_2(4, 48) = 1.14$, $MSE = 2,005$, $p = .35$. Post hoc analyses of the condition effect (Newman–Keuls test with $p < .05$) revealed that all three experimental conditions differed significantly from each other in the participant analysis. In the item analysis, the difference between the homophones and the HF controls failed to reach significance.

For the error rates, we found a main effect of condition (12.5%, 9.4%, and 3.7% for the LF-control, homophone, and HF-control conditions, respectively), $F_1(2, 42) = 9.75$, $MSE = 1.08$, $p < .001$; $F_2(2, 24) = 4.25$, $MSE = 6.07$, $p < .05$.
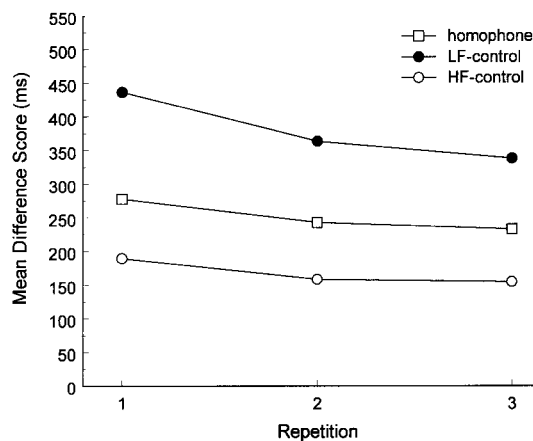

Figure 2. Mean difference scores (in ms) by condition and repetition from Experiment 2.

Post hoc analyses revealed that the LF controls differed from the HF controls in the participant analysis and in the item analysis. The difference between the homophones and the HF controls was only significant in the participant analysis. Error rates decreased with repetition, $F_1(2, 42) = 29.42$, $MSE = 0.63$, $p < .001$; $F_2(2, 48) = 34.00$, $MSE = 1.34$, $p < .001$. There was a trend toward smaller condition differences over repetitions, but the interaction was only marginally significant, $F_1(4, 84) = 2.38$, $MSE = 0.50$, $p = .06$; $F_2(4, 48) = 2.16$, $MSE = 1.34$, $p = .09$.

Overall, the data pattern is similar to the pattern obtained by Jescheniak and Levelt (1994, Experiment 6) and in Experiment 1 of the present article, providing a cross-linguistic replication of the basic finding. In all cases, the (low-frequency) homophones were named significantly faster than the LF controls, and this difference did not dissipate over repetitions.

General Discussion

The results of Experiments 1 and 2 confirm our prediction that the existence of a high-frequency homophone affects the speed with which Dutch and German target words are produced in response to English probes. In both experiments, low-frequency target words with a high-frequency homophone twin were produced substantially faster than low-frequency control words without a high-frequency twin. This pattern strongly suggests that homophonic twins share a representation in the lexical system, most likely the word form representation.

Despite the similarities between the results of the Dutch and the results of the German experiments, one needs to acknowledge the differences in detail. In contrast to the original Dutch experiment (Jescheniak & Levelt, 1994, Experiment 6) and its replication (present Experiment 1), a reliable difference between the HF control and the homophone condition was obtained in the German experiment. The homophone effect amounted to only 60.8% of the frequency effect in the German experiment, compared with 83.1% in the Dutch experiment. At present, we can only speculate why this was the case. To start with, the selection of the German experimental materials was heavily constrained, and the item set was small. As mentioned earlier, the set of HF controls could not be perfectly matched with the sum frequency of the homophones and the high-frequency twins. This might have contributed to the observed difference. Another possible explanation is that translation agreement was lower for the homophones than for the LF controls and the HF controls. To explore this possibility, we asked an independent group of 27 participants (recruited according to the same criteria as the participants from Experiment 2) first to write down the best German translation for each of the English probes and then to judge the appropriateness of the German translation targets used in the experiment. The results of these translation agreement tests are presented in Table 3.

There were three notable findings. First, translation appropriateness ratings were equally high for the homophone set and the LF-control set, $t_1(25) = 1.43$, $p = .17$; $t_2(16) = 1.11$, $p = .28$. Second, the proportion of trials on which the words we used as targets in the experiment were generated spontaneously did not differ between homophones and LF-controls, $t_1(26) = 0.72$, $p =$

Table 3
Results From the Translation Agreement Test

| | Condition | | | | | |
| Agreement test | Homophones | | LF controls | | HF controls | |
	M	SE	M	SE	M	SE
Appropriateness rating	4.86	0.17	5.19	0.22	6.46	0.07
% target responses	3.7	1.4	4.9	1.4	68.0	2.7
% valid other responses	50.0	2.7	11.0	1.9	15.0	1.8

Note. A 7-point scale was used for collecting appropriateness ratings (1 denoting *inappropriate translations* and 7 denoting *best translations*). LF = low frequency; HF = high frequency.

.48; $t_2(16) = 0.45$, $p = .66$. These values were low, but in the main experiment, participants were trained to use the target translations, and the participants actually produced them on the vast majority of trials without error. Third, valid nontarget responses (trials in which participants responded, for instance, with *Feier* instead of the expected *Fest* to the probe *celebration*) occurred on a much larger proportion of trials for the homophones than for the LF controls, $t_1(26) = 16.83$, $p < .001$; $t_2(16) = 2.81$, $p < .05$. This suggests that more nontarget lexical candidates may have been activated during the translation task for the homophones than for the LF controls. These competitors could have interfered with the processing of the target more strongly in the homophone condition than in the LF-control condition, leading to the observed attenuation of the homophone effect. Other explanations are also possible and might be explored in future experimental work. For the time being, the most important conclusions are that performance for homophones differed from performance for LF controls and that this pattern was obtained with different sets of materials and in different languages.

Additional chronometric evidence in favor of shared representations of homophones comes from a recent study by Schriefers, Jescheniak, and Hantsch (2002), investigating the production of German noun phrases. In this study, participants produced singular noun phrases (e.g., *das Haus*–the house) or plural noun phrases (e.g., *die Häuser*–the houses) in response to one or two target objects. When the singular and plural determiners had different forms, plural noun phrases were produced more slowly than singular noun phrases. This was the case for object names of masculine and neuter gender in which the singular determiners are *der* (masculine) and *das* (neuter), and the plural determiner is *die*. By contrast, when the singular and plural determiner were homophones, plural noun phrases were produced slightly faster than singular noun phrases. This was the case for object names of feminine gender in which the singular and plural determiner is *die*. This pattern suggests two things. First, the number effect for nouns of masculine and neuter gender indicates that in the production of plural noun phrases, singular and plural determiners were activated and competed with each other. Second, and most important for the issue at hand, the absence of a number effect for nouns of feminine gender in plural noun phrase production suggests that the grammatical features *singular* and *plural* activated the same form representation, shared by the singular and plural feminine definite determiner lemmas. Note that the same pattern of results has been

obtained by Janssen and Caramazza in Dutch, which has a deter-miner system with similar properties as German (reported in Caramazza, Miozzo, Costa, Schiller, & Alario, 2001). These chronometric findings clearly demonstrate the existence of overlapping lexical representations also for homophonic closed class elements, which is in line with Dell's (1990) speech error data and in contrast to what Caramazza, Costa, et al. (2001) considered a possible source for the difference between their own findings and Dell's earlier observations.

In summary, the results of the new translation experiments reported in this article and the noun phrase production experiments by Schriefers, Jescheniak, and Hantsch (2002) and Janssen and Caramazza (reported in Caramazza, Miozzo, et al., 2001) provide converging cross-linguistic evidence consistent with the view that homophones have distinct lemmas but share a word form representation, as suggested, for instance, by Dell (1990), Jescheniak and Levelt (1994), Cutting and Ferreira (1999), Levelt et al. (1999), and Ferreira and Griffin (in press). We find it difficult to reconcile these findings with the proposal of fully independent representations for homophones and the absence of a lemma level.

Still, the Spanish–English translation experiment by Caramazza, Costa, et al. (2001; their Experiment 3A) did not show any frequency inheritance by low-frequency homophones. In addition, Bonin and Fayol (2002) failed to obtain evidence for full frequency inheritance between homophones in spoken and written picture-naming experiments carried out in French. Bonin and Fayol used pairs of pictures with homophonic names differing in spelling and frequency, for example, *lait* (milk) and *laie* (wild sow). The participants named the high-frequency members of the pairs considerably faster (on average by about 90 ms) than the low-frequency members, which is unexpected if there was full frequency inheritance. However, because the experiment did not include a low-frequency control condition, it is unknown whether there was partial frequency inheritance. Curiously, a large reversed frequency effect (of 300 ms) was found in a speeded semantic judgment task on the same items. Thus, again the high- and low-frequency items may not have been optimally matched in all relevant aspects (see also Osborne & Burke, 2002, for evidence that differences in the ease of stimulus recognition might mask frequency-inheritance effects in a picture naming task).

Clearly, further research is needed to determine under which conditions evidence for frequency inheritance between homophones can be obtained. The available experiments differed in the tasks used to elicit the target names and in the control tasks as well as in the types of homophones that were studied. Some experiments used homographic homophones; others used heterographic ones or a mixture of both types of homophones. The experiments also differed in whether or not the homophones belonged to the same syntactic class and in whether they were related or unrelated in meaning. For instance, in Bonin and Fayol's (2002) study, the pairs were heterographic; they were all nouns and unrelated in meaning. In Experiment 2 of the present study, the members of the pairs were homographic; they differed in syntactic class, and most of them were unrelated in meaning.

In addition, the existing experiments differed in the languages that were studied. Perhaps frequency inheritance in homophones is more likely to be observed in some languages than in others. The reason may be that speakers of different languages retrieve words in slightly different ways, allowing, for instance, more or less time for monitoring or verification processes. Dutch and German have relatively few homophones and revealed frequency inheritance effects (Jescheniak & Levelt, 1994, and the present study). Chinese has relatively many homophones but did not show an inheritance effect (Caramazza, Costa, et al., 2001). English also has many homophones, but here the situation with respect to frequency inheritance is less clear. Caramazza, Costa, et al. (2001) did not find such an effect, but Dell (1990) did. Evidently, careful cross-linguistic comparisons are needed to assess the issue of possible language-specific differences in homophone inheritance and to provide a coherent theoretical account. For now, we conclude that, in view of the inheritance patterns found in Dutch and German, the distinction of lemma and word form representations continues to be useful when theorizing about lexical representations accessed during speech production.

References

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

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

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

Caramazza, A., & Miozzo, M. (1998). More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition, 69,* 231–241.

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F. -X. (2001). A cross-linguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 209–226). Cambridge, MA: MIT Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Cutting, C. J., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 318–344.

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

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42,* 287–314.

Ferreira, V. S., & Griffin, Z. M. (in press). Phonological influences on lexical (mis-) selection. *Psychological Science.*

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *The Cambridge survey of linguistics. Vol. 3. Language: Psychological and biological aspects* (pp. 69–96). Cambridge, MA: Harvard University 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.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Science, 3,* 223–232.

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.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Osborne, G. L., & Burke, D. M. (2002). Frequency and aging effects on naming pictures of homophones. *Abstracts of the Psychonomic Society, 7,* 97.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition, 69,* 219–230.

Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 941–950.

Appendix

Materials Used in Experiment 2

German Target Words and English Probe Words (in brackets) by Experimental Condition (Filler Items Included)

Homophones	LF Controls	HF Controls	Animate Filler Items		
All [universe]	*Ast* [bough]	*Art* [manner]	*Adler* [eagle]	*Hai* [shark]	*Meise* [titmouse]
Alt [counter-tenor]	*Beil* [hatchet]	*Bild* [picture]	*Affe* [monkey]	*Hecht* [pike]	*Nixe* [mermaid]
Fest [celebration]	*Brett* [plank]	*Fall* [case]	*Bote* [messenger]	*Hengst* [stallion]	*Oma* [grandmother]
Gut [estate]	*Erz* [ore]	*Jahr* [year]	*Braut* [fiancee]	*Herr* [gentleman]	*Pferd* [horse]
Hoch [cheer]	*Fach* [compartment]	*Tag* [day]	*Dieb* [thief]	*Hirte* [shepherd]	*Pute* [turkey-hen]
Laut [sound]	*Guß* [founding]	*Teil* [component]	*Dohle* [jackdaw]	*Igel* [hedgehog]	*Specht* [woodpecker]
Tief [depression]	*Hof* [courtyard]	*Weg* [path]	*Esel* [donkey]	*Kind* [child]	*Strauß* [ostrich]
Trotz [defiance]	*Hohn* [derision]	*Zeit* [time]	*Fee* [fairy]	*Lachs* [salmon]	*Wirt* [host]
Wohl [welfare]	*Wall* [rampart]	*Ziel* [destination]	*Hahn* [rooster]	*Maurer* [bricklayer]	*Zwerg* [dwarf]

Note. LF = low frequency; HF = high frequency.