

Prosody and lemma selection

CONRAD PERRY and JIE ZHUANG
University of Hong Kong, Hong Kong, China

In three picture-naming experiments, we examined the effect of prosodic context on the synonyms people use to name pictures in Mandarin Chinese. This was done without time pressure. The results showed that when monosyllabic and bisyllabic synonyms (e.g., *hen/chicken*) were embedded in a context of pictures with either bisyllabic or trisyllabic names, participants gave bisyllabic responses to the synonyms more often than they did in a condition without such a context. The difference was very similar in magnitude in both the bisyllabic and trisyllabic contextual conditions. These results suggest that people are biased toward using synonyms that have numbers of syllables equal or similar to those of the prosodic context. If it is assumed that prosodic effects originate at a stage of processing beyond the lemma level, then this suggests either that multiple phonological forms of synonyms can be activated or that there is feedback from prosodic processing that influences lemma selection.

One of the interesting questions in spoken word production is where the selection of a single response for a given concept is made. In the model of Levelt, Roelofs, and Meyer (1999), which proposes a distinction between a lemma level (which includes the syntactic properties of words) and a form level (which includes the phonological characteristics of words), selection is assumed to occur at the lemma level. Such a selection procedure means that only one phonological word form is ever active at a given time in the model.

Whether more than one phonological word form can actually be activated at a given time remains under dispute. Peterson and Savoy (1998; see also Cutting & Ferreira, 1999, and Jescheniak & Schriefers, 1998) provided evidence that multiple lemmas are selected in speech production and, hence, can cause multiple phonological forms to be active. They examined phonological priming using pictures that typically elicit two different names—that is, synonyms (e.g., a picture of a large cushioned chair produced the answers “couch” and “sofa”) and for which one of the names was dominant (e.g., “couch” was given by 80% of the participants and “sofa” by only 20%). The results that they found showed that early in the time course of processing, both phonologies (even that of the word form that was not typically selected) could be primed to similar extents. Later in the time course of processing, however, only the dominant word form could be primed. Peterson and Savoy suggested that this showed that two phonological forms were activated by the synonyms early in spoken word production but that the phonology of the

dominant form inhibited the phonology of the other later in processing.

The findings of Peterson and Savoy (1998) were problematic for theories of spoken word production that suggest that individual lemmas are selected before phonology is generated. This is because the data suggest that a single picture can activate the phonology of two lemmas and that when this occurs, the activation of one of the word forms is reduced later in processing due to cascaded activation. To account for this result with a single lemma selection process, Levelt et al. (1999) suggested that when lemmas are activated to very similar extents—such as when pictures activate near-synonyms—more than one lemma may be selected, particularly under time pressure. On the basis of the idea that time pressure may cause unusual behavior, Levelt et al. were able to adhere to their general principle that no more than one lemma is selected at a time in normal speech.

Common tasks used to investigate lemma selection typically require participants to name pictures under time pressure. This poses a problem for testing the possibility that only one lemma tends to be activated in relaxed naming conditions even when two are activated in speeded naming conditions. Thus, an experimental paradigm that can be used to examine whether multiple phonological forms are activated even without time pressure is needed.

A potential way to examine whether phonological effects can influence lemma selection under relaxed conditions is to create a prosodic context that might bias people in their choice of lemmas (see Rapp & Samuel, 2002, for another method that could be used without time pressure). The idea is that if a single lemma is always selected before its phonology is generated and phonology cannot feed back and influence lemma selection, then phonological contexts would be expected not to bias lemma selection. In this case, because the phonological form can be biased only after selection, only aspects of

C.P. was supported by a UDF grant from the University of Hong Kong. We would like to express our thanks to Yang Jing and Ning Ning for collecting the data and to Steve Mathews and Richard Wong for helpful discussions about Chinese. Correspondence concerning this article should be addressed to C. Perry (e-mail: perry@up.univ-mrs.fr).

the phonology of the lemma, rather than lemma selection itself, should be subject to biasing.

It is possible to create a prosodic context by using filler words that have a given nonsemantic and nonsyntactic contextual property. In particular, different prosodic contexts can be created by using filler pictures that can be named with words having a set number of syllables. The effect that these fillers have on the naming of synonyms can then be examined. The idea behind this is that if people habituate to a prosodic context with a set number of syllables, they may be more likely to select a lemma that has the same (or a similar) number of syllables. For instance, if people hear a list of bisyllabic words (e.g., *lagging, paper, sofa*), when ambiguity exists they may be more likely to use a bisyllabic than a monosyllabic name (i.e., they may be more likely to say “chicken” than “hen”) than they would be in a situation without such a context.

Although it might be possible, we weren't successful at finding a set of pictures in English to which people commonly assign two different names that differ in syllable length (e.g., *hen/chicken*). In general, it appears that in English synonyms tend to have the same number of syllables. Unlike in English, however, in Mandarin Chinese there are many synonyms that have different numbers of syllables. To some extent, this is due to the historical evolution of the language, whereby words that were initially monosyllabic slowly became bisyllabic for a number of different reasons (see, e.g., Feng, 1998). However, the monosyllabic versions of such words did not always disappear; rather, for many words both the bisyllabic and monosyllabic versions are commonly used. There are a number of other potential reasons for which one- and two-syllable synonyms occur in Chinese, and Duanmu (1999) provides a good summary of the various theories.

In the experiments reported below, we examined whether prosodic context could influence lemma selection. In the first and second experiments, we examined whether a bisyllabic prosodic context would bias people toward using bisyllabic picture names when ambiguity exists (i.e., when there are synonyms). In the third experiment, we examined whether a trisyllabic prosodic context would bias people toward using bisyllabic names. The goal was to see whether, to cause an effect, there needed to be an exact match between the number of syllables used in the prosodic context and the number of syllables in potential synonyms, or whether biasing could be induced simply by pictures whose names had more syllables than their potential synonyms did.

EXPERIMENT 1

Method

Participants. Twenty students at Beijing Normal University participated in the experiment for a small monetary remuneration. All of the participants were native speakers of Mandarin Chinese.

Stimuli. Forty pictures were selected for which two common near-synonyms could be used. An additional 66 pictures, each of which typically had only a bisyllabic name, were used as fillers. In

addition, 25 pictures were used as practice stimuli before the task began. Twelve of these had monosyllabic names, and 13 had bisyllabic names. All were hand-drawn black-and-white line drawings. A full list of the synonyms used, including the monosyllabic and bisyllabic names that were expected, appears in the Appendix.

Procedure. The participants were asked to name each picture in the set with the first name they could think of. They were told that speed was not important, but rather that they should try not to make errors in the task.

The stimuli were arranged in two counterbalanced groups. In one group, half of the pictures with two nearly synonymous names appeared in a block at the start of the task (neutral context condition). This was followed by 15 filler pictures, which had only bisyllabic names. The other half of the pictures with two nearly synonymous names then appeared, randomly distributed within 51 filler pictures that had only bisyllabic names (bisyllabic context condition). In the other counterbalanced group, the synonym pictures that appeared in the neutral context condition appeared in the bisyllabic context condition, and vice versa. Thus, the stimuli used in the first counterbalanced group were presented in the following order: (1) practice examples, (2) synonym-only block, (3) bisyllabic-picture-only block, (4) synonym and bisyllabic picture block. The stimuli in the second counterbalanced group were presented in identical fashion, except that the order of presentation of (2) and (4) was reversed.

In individual trials, each picture appeared on the screen until the participant made a response. Once a response had been made, the picture disappeared. After 800 msec, the next picture appeared. No feedback was given in the task.

Results

The responses of 1 participant, which we had difficulty interpreting, were completely removed from the analysis. Four items for which the participants most often used names other than the two expected synonyms were also removed from the analysis. An additional 4.7% of the individual responses that were not one of the two expected synonyms were removed. Mean response probabilities appear in the Appendix. The mean speed at which the participants responded to the pictures was calculated using the response times (RTs) from both the filler and the critical pictures. RTs that took longer than 2 sec were replaced with 2-sec RTs so that particularly slow individual responses would not bias the overall mean. There were 203 of these.

The results showed that in the neutral context the participants gave monosyllabic answers 55.3% of the time, and in the bisyllabic context they gave monosyllabic answers only 43.5% of the time. That difference was significant [$t_1(18) = 3.09$, $SD = 16.4$, $p < .01$; $t_2(35) = 3.15$, $SD = 22.4$, $p < .005$]. RTs appeared to be relatively long in comparison with RTs found in typical picture-naming tasks, with the mean RT to each picture being 1,230 msec ($SD = 144$ msec). (By way of comparison, participants from a similar pool averaged around 750 msec in a speeded picture-naming task in the study of Caramazza, Costa, Miozzo, & Bi, 2001.)

The pattern of results was clear. The participants were less likely to use monosyllabic synonyms (and hence more likely to use bisyllabic synonyms) when the pictures were embedded in a bisyllabic context than when they were presented in a neutral context. This suggests that contextual biasing caused by processes that occur

after lemma activation (i.e., nonsemantic and nonsyntactic processes) can be present.

EXPERIMENT 2

The previous experiment suggested that lemma selection could be biased by a context created by varying the number of syllables in filler items. One potential influence on the results is that people might be susceptible to giving more monosyllabic answers early in a list than late in a list, when bisyllabic fillers are used. One reason for this is that although only 26.7% of the 9,000 most frequently used Chinese words are monosyllabic (Beijing Language Institute, 1986), monosyllabic forms are overrepresented among high-frequency words (a similar phenomenon occurs in English). Thus, the presence of monosyllabic words as a proportion of all words in normal speech is likely to be greater than their presence in the bisyllabic filler condition in our list. Therefore, if there is an initial bias toward words with a number of syllables typical of words used in normal speech, it is likely to be present at the start of a list in comparison with after the presentation of many bisyllabic fillers.

Because the proportion of the entire item list that is used for the bisyllabic context condition is greater than that which is used for the neutral context condition, it is not possible to get around this problem by using a counterbalanced design in which half the participants get the critical pictures in the bisyllabic context first and the other half get the critical pictures in that context second. Thus, if more monosyllabic answers are given at the start of a list for only a small number of items, then there would still be a bias toward giving monosyllabic words in the group that began the task with the critical items in a neutral context. This is because many fillers would be present at the start of the task, when the critical items are in a context of bisyllabic words as opposed to no context, as in the neutral condition; hence, any early bias on critical items is reduced for the bisyllabic context group in comparison with the neutral context group.

To investigate whether or not the order of presentation confounded our results in the previous experiment, we ran an experiment that was identical except that (1) all of the practice items had bisyllabic names, (2) the critical items were first presented in the bisyllabic context, and (3) the filler pictures that occurred between the two blocks of critical stimuli had monosyllabic names. In addition, since the task was not speeded, at the end of the task we also asked the participants whether they noticed any patterns, with respect to word length, in the way the pictures had been presented.

Method

Participants. Twenty-six students from Beijing Normal University participated in the experiment for a small monetary remuneration.

Stimuli. The same critical stimuli that were used in Experiment 1 were used in the present experiment. The 15 filler items that appeared between the two context conditions in the previous experiment were replaced by pictures that could have only monosyllabic names.

Procedure. The procedure was identical to that of the previous experiment except that (1) the participants were queried after the experiment about whether they had noticed any pattern in the way the pictures were presented, with respect to word length; and (2) the order in which the stimuli were presented was reversed in terms of list context. That is, the critical stimuli were first presented in a bisyllabic context.

Results

On debriefing, no participant reported any information that led us to suspect that he or she might have had an idea about the design of the task. Of the individual data, 6.1% of the individual responses that were not one of the two expected synonyms were removed. In terms of the RT analysis, 795 RTs longer than 2 sec were replaced with 2-sec RTs. Twenty-one items with RTs shorter than 100 msec were removed from the analysis.

The results showed that in the neutral context the participants gave monosyllabic answers 51.9% of the time, and in the bisyllabic context they gave monosyllabic answers only 39.6% of the time. That difference was significant [$t_1(25) = 3.19$, $SD = 19.9$, $p < .005$; $t_2(35) = 3.96$, $SD = 25.5$, $p < .001$]. The mean RTs to the pictures were again rather long, with a participant average of 1,098 msec ($SD = 198$ msec).

The pattern of results was clear. The participants were less likely to use monosyllabic synonyms (and hence more likely to use bisyllabic synonyms) when the words were embedded in a bisyllabic context than when they were presented in a neutral context. The absolute magnitude of the effect was almost identical to that of the previous experiment. It therefore appears that whether the critical items are presented in a bisyllabic context at the start of an experiment or at its end makes very little difference in terms of the pattern of responses. In addition, since no participant reported any information about the order of the items in the task, it seems unlikely that these results were contaminated by the participants' guessing what the design of the task was and then responding in an atypical way.

EXPERIMENT 3

The previous two experiments demonstrated that lemma selection could be biased by the prosodic context of the list. The most obvious interpretation of this is that people prefer to use picture names that have the same number of syllables as the name of the context. Such a possibility is supported by a number of implicit priming experiments (Roelofs & Meyer, 1998). In those experiments, it was shown that people could recall the second word of an associated word pair faster if they could predict the number of syllables it had and its first syllable, in comparison with when they could not. When they could predict the first syllable but not the number of syllables, RTs were not decreased. The results were interpreted as suggesting that people are able to plan the metrical structure of a word in advance if the specific metrical structure (including the number of syllables) is also known. A similar interpretation could be used for the

previous experiment, in which it would be assumed that people might be able to generate a metrical structure before seeing the stimuli in the bisyllabic context condition, or at least that they exhibit a bias toward lemmas that fit an expected metrical structure.

Although the metrical structure of the individual lemma may be one locus of the context effect, the biasing may also be due to other aspects of spoken word production, such as rhythm (see, e.g., Hayes, 1995) or the timing assigned to a word via a prosody generator (Ferreira, 1993). Thus, it may be that the effects found in the experiments are not due to the exact metrical structure of the picture names, but rather to the possibility that people are biased by other aspects of speech production, and one of those preferences may be related to aspects of speech timing. This would be particularly interesting, since such aspects of spoken word production are typically presumed to come at a later stage of processing than the retrieval of metrical information for an individual word.

Perhaps a likely locus of such an effect would be the timing interval assigned to each word in a sentence. In our case, this corresponds to the timing interval assigned to the word for each picture. At least according to Ferreira (1993), such timing intervals are assigned by a prosody generator as slots into which segmental phonology is inserted (see also Meyer, 1994, for other possible explanations). Since the duration of such timing intervals must be extremely flexible to cope with different pragmatic contexts (e.g., speaking at different rates), tasks that emphasize longer words might also increase the typical time interval assigned by the prosody generator. If the length of the timing interval then has some effect on which lemma is selected when conflicts exist, it may cause a bias toward using synonyms with different lengths. An alternative to this explanation is that people like to speak with regularly timed stressed vowels (see Meyer, 1994, for a discussion). In this case, instead of initially coming from the optimal fitting of a lemma's metrical information into timing slots, the effect might come from metrical information about lemmas that allows the most typical vowel rhythm to be used. Unfortunately, the data reported here do not allow any sort of differentiation between these two possibilities.

One point that we should note is that Mandarin Chinese might be particularly sensitive to timing effects, since it has been argued (Duanmu, 1999; Feng, 2003) that the grammaticality of some sentence types is determined by the number of syllables a particular word has. Thus, certain types of sentences can be constructed grammatically only by choosing words on the basis of both their prosodic and their syntactic characteristics, rather than just on the basis of the latter. Duanmu offers this as one reason for which so many singular concepts are named with both monosyllabic and bisyllabic synonyms in Chinese. He suggests that this reflects a historical tendency in word creation that is caused by a need to construct simple sentences that are grammatical, or at least grammatically preferable, on the basis of their metrical characteristics, which he argues is important in Chinese.

By changing the context words that are used, it is possible to investigate whether the biasing effect is specific to the bisyllabic nature of the names of the context pictures used or is due to a more general increase in the number of syllables in the context words. In this experiment, a context of pictures with trisyllabic rather than bisyllabic names was used. If there is an effect similar in size to the effect of the previous experiment, this would suggest that the effect comes from a process related to more than the ability to predict the metrical frame of a word. Alternatively, if there is no effect, this would suggest that it is necessary to use pictures with names having an identical number of syllables in order to find such a bias.

Method

Participants. Twenty students from the Beijing Normal University participated in the study. All were native speakers of Mandarin Chinese.

Stimuli. The critical pictures were the same as those used in Experiment 1. However, each filler picture used to create the contextual conditions corresponded to a picture that had only a trisyllabic name. Since it was not possible to find enough pictures that had a single trisyllabic name, only 33 were selected. They were each repeated, for a total of 66 filler words.

Procedure. The procedure was the same as that of Experiment 1, apart from the repetition of the context words.

Results

Three pictures for which the participants gave responses other than one of the two expected synonyms more than 50% of the time were removed from the analysis. An additional 10.7% of the individual items were removed for the same reason. In terms of the RT analysis, 186 RTs that were over 2 sec long were replaced with 2-sec RTs. An additional 17 items that triggered the voice key in less than 100 msec were removed from the analysis.

The results were very similar to those of the previous experiment. In the neutral context condition, the participants used monosyllabic answers 57.9% of the time, and in the trisyllabic context condition they used monosyllabic answers 46.9% of the time. That difference was significant [$t_1(19) = 2.45, SD = 19.3, p < .05; t_2(36) = 2.65, SD = 19.7, p < .05$]. Responses were quite slow in comparison with those found in typical speeded picture naming tasks, with an average RT per participant of 1,023 msec ($SD = 74$ msec).

The results showed that using a context condition consisting of pictures with trisyllabic names was sufficient to bias the participants toward using bisyllabic rather than monosyllabic names. If metrical frames store the exact number of syllables a word has, and if even partially overlapping frames are completely independent of each other, then this suggests that the bias on lemma selection must have originated at a stage of processing that occurs after the processing of individual word forms.

POST HOC ANALYSIS: CHINESE LEMMAS

Although the three experiments displayed an effect of biasing from list context, one potential argument against

the effect's occurring at a lemma selection level is that the participants may have used some form of postlemma searching strategy, since they did not have to respond to the pictures quickly. The participants may have (1) monitored their own speech, (2) tried to search through all potential synonyms that they could generate in a serial manner, and (3) on the basis of the list of lemmas they retrieved serially, tried to produce an answer most congruent with the list conditions. This must have been done without their conscious awareness, since none reported any awareness of the strategic arrangement of the items in Experiment 2 when we explicitly questioned them after the task.

One way to examine this hypothesis would be to run the task under speeded conditions and see if the same results were found. However, the main goal of these experiments was to demonstrate synonym effects without time pressure. Therefore, a method of assessing whether a strategic bias was occurring due to a lemma-searching strategy without time pressure is necessary.

One such method is to examine pairs of synonyms that have different properties and to show that one type of synonym pair is open to biasing in the task but another is not, even though the retrieval of both members of both pairs should not be difficult. This would show that people do not simply generate all possible pronunciations for names of a given picture and then choose one. If that were the case, synonyms represented differently should not differ from each other in terms of susceptibility to bias.

It is possible to do this on the basis of the groups that we used. In our stimulus set, 24 of the synonyms for the pictures were bisyllabic and did not share the first syllable with its monosyllabic counterpart (e.g., 烟 and 香烟), whereas 16 did share the first syllable (e.g., 鸭子 and 鸭). Depending on the assumptions made about what constitutes a word in the Chinese lexicon (see Packard, 2000, for a review of different perspectives on how words might be represented in the Chinese lexicon), the stimuli may have different types of representations. At least in the first case, the situation seems quite similar to that of languages such as English (e.g., *gun* and *handgun*), in which it is not typically believed that both words would be represented by one lemma. This is because, although the words are semantically related and even share a morpheme, the first morpheme differs in the two words and appears to be critical in specifying the word (see Taft & Forster, 1976, for experimental evidence of this).

Alternatively, the second type of synonym might be represented by one lemma, and the second syllable may represent an optional syllable that can be used depending on pragmatic situations. This is particularly the case for our stimuli, because the first syllables generally carried a meaning congruent with the picture whereas the second syllables often had no semantic relationship to the words at all. Thus, for example, the morphemic meaning of 子 has essentially nothing to do with "duckiness" apart from the fact it happens to be used in 鸭子 (duck).

This is because it is a syllable used to make what was once a monosyllabic word bisyllabic (a word-forming affix; Packard, 2000). Similar situations occur in English. For example, in some English dialects a common phonological marker can be put on the end of some names to form a nickname (e.g., *Frank-y*, *Dave-y*, *John-y*), and whether people use one form or the other depends on pragmatic considerations. At least in this example, in formal situations it is unlikely that one would use the nickname instead of the formal name (e.g., "Good day, Mr. Boss, my name is Johnny"), whereas in informal situations one might frequently use a nickname (e.g., "Hi, pal, my name's Johnny"). A similar situation occurs in adult and child speech. By way of example, most people know that the word *fishies* is a near-synonym of *fishes*, even though one is colloquial and the other not.

If we accept that both members of such synonym pairs are represented by a single lemma that contains the base morpheme and a list of morphological alternatives, then people must have some way of distinguishing when to use one and not the other. For our English examples, this could be done via a diacritic parameter that specifies which version of a word to use in formal situations. This would suggest that this type of synonym pair might be represented quite differently from those that do not have the same first morpheme.

If bisyllabic synonyms with a shared first morpheme are represented differently than bisyllabic synonyms that do not share a first morpheme, then the effect of prosody on them may be quite different in comparison with its effect on synonyms with two lemmas. This is because if prosody influences lemma selection on synonyms with only one lemma, then it would need to have an influence on the diacritic parameter; thus, the bisyllabic form would be chosen more often in context than in a selection process between two different phonologies generated by two different lemmas. Thus, if (1) the phonology of more than one lemma is generated automatically with pictures that are represented by two lemmas and (2) reading out the names of a series of pictures with different numbers of syllables is not conducive to modifying diacritic parameters in such a way that exactly the same sized effect is found as that due to the influence of speech timing on the selection of multiple phonologies, then this suggests that prosody might have a different effect on the two different types of synonyms. Alternatively, if the results of our previous experiments are due to a strategic speech-monitoring bias by which people choose synonym names most congruent with the list, then both types of synonym should be biased. This is because there would be no reason to assume that recalling both near-synonyms associated with only one lemma would be more difficult than recalling a pair of synonyms associated with two different lemmas.

To investigate whether there was a difference in the extent of the effect of list context on the two different types of synonyms used in the experiments, we collapsed the results from all three experiments and then split the

results into groups on the basis of whether the bisyllabic synonym in the pair shared the same first morpheme as the monosyllabic synonym. This meant that there were responses to 46 pictures that could potentially be named using two synonyms that shared the first morpheme and responses to 67 that did not. We then performed a 2×2 analysis of variance (ANOVA) on list context (multisyllabic vs. neutral) and type of synonym. The results showed that there was a significant main effect of synonym type [$F_1(1,64) = 115.45$, $MS_e = 206$, $p < .001$; $F_2(1,111) = 12.82$, $MS_e = 1,924$, $p < .005$], with more monosyllabic answers given for the group of synonym pairs that did not share a first morpheme than for the group that did, and a significant main effect of list context [$F_1(1,64) = 24.53$, $MS_e = 304$, $p < .001$; $F_2(1,111) = 26.11$, $MS_e = 251$, $p < .001$], with more bisyllabic answers given for multisyllabic list contexts. The interaction was also significant [$F_1(1,64) = 5.84$, $MS_e = 340.28$, $p < .05$; $F_2(1,111) = 10.79$, $MS_e = 231$, $p < .005$]. The interaction appeared to be caused by the fact that the synonyms that did not share the same first morpheme showed a much larger biasing effect (17.2%) than the synonyms that did (3.7%). Further post hoc tests on the individual groups showed that the synonyms that did not share the same first morpheme were significantly affected by the biasing context [$t_1(64) = 4.79$, $SD = 27.4$, $p < .001$; $t_2(66) = 5.60$, $SD = 22.2$, $p < .001$], whereas those that shared the same first morpheme were not [$t_1(64) = 1.80$, $SD = 23.24$, $p = .077$; $t_2(45) = 1.78$, $SD = 14.3$, $p = .082$]. These results appear in Table 1.¹

The results of our analysis of the biasing effect with respect to the different types of synonyms go against the possibility that they were due to the participants' unconsciously going through a list of all easily retrievable synonyms serially and then choosing the one that was most congruent with the list context. If they had, then both types of synonyms should have been biased by list context. However, that was not the case: Synonyms that shared the same first morpheme showed very little effect of biasing. Of course, it would always be possible to argue that people recall all potential synonym names serially but that, when they recall two names on the basis of a diacritic parameter, they do not allow prosody to influence the decision whereas, when they recall two names based on two different lemmas, they do. However, this is essentially an ad hoc suggestion, and we can think of no obvious reason to believe it to be true.

Table 1
Percentage of Monosyllabic Answers Given Across the Three Experiments as a Function of Synonym Type and List Context

Synonym Type	Percentage of Monosyllabic Answers		
	List Context		
	Neutral	Multisyllabic	Difference
Share first morpheme	38	34	4
Do not share first morpheme	66	49	17

DISCUSSION

The extent to which multiple phonological forms are activated in speech production has been an issue of recent interest. Some evidence has suggested that under time duress people can activate more than a single phonological form for a single concept (see, e.g., Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998). One argument that detracts from these results is that such activation occurs not because it is typical of speech production but rather because of the emphasis on speed in some experimental tasks (Levelt et al., 1999). Therefore, in this study we tried to find evidence that more than one phonological form can be activated without time pressure.

To try to find such evidence, we examined the number of syllables in synonyms people use in different prosodic contexts. The idea behind this is that prosodic effects occur after lemma selection in some speech production models. In the models of Levelt et al. (1999) and Dell (see, e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), for instance, it is assumed that people don't know the number of syllables the name of a concept has until after a lemma has been activated. Therefore, if prosodic effects influence the selection of a name for a concept, it may be that (1) either activation from prosodic processing can feed back and influence lemma selection or (2) multiple phonological forms are selected, and that the selection of these forms is influenced by prosodic conditions.

In our first and second experiments, we examined whether a bisyllabic picture context would bias people toward using bisyllabic names when one synonym was monosyllabic and the other bisyllabic. This was done by embedding pictures that each had a monosyllabic and a bisyllabic name that were nearly synonymous within a group of pictures that had only bisyllabic names, and comparing the number of bisyllabic responses people gave to the same synonyms when they were not embedded in such a context. The results showed that people preferred to use bisyllabic names in a bisyllabic context. In the third experiment, we examined whether this bias was restricted to a particular prosodic condition (i.e., bisyllabic words) or was more generally related to the length of the context words. To create such a condition, we examined the effect of a trisyllabic context on the choice of a synonym. The results were essentially the same, with people choosing more bisyllabic than monosyllabic words in the trisyllabic context.

The results of the three experiments suggest that it is possible to find phonological effects on lemma selection even without time pressure. Furthermore, they suggest that the biasing emerges from a general number-of-syllables constraint, since both bisyllabic and trisyllabic contextual conditions caused effects of very similar magnitudes. This is interesting, since it suggests that individual lemma selection is influenced by a rather general

prosodic property rather than by a property specific to individual phonological word forms, such as the number of syllables in a word. By “general prosodic property,” we mean a prosodic property that might be generated by an independent prosody generator, such as an intonation pattern (see, e.g., Hayes, 1995) or the expected duration of a word in a sentence (see, e.g., Ferreira, 1993), rather than by something associated with a specific lexical item, such as which syllable in a word should be stressed. Thus, in models of speech production in which properties of individual words are first retrieved and later modified for more general purposes such as those involved in sentence production (e.g., Ferreira, 1993), it would have to be assumed that either multiple levels of feedback can occur, starting at the time when prosodic processing due to the prosody generator occurs, or that multiple phonological word forms are activated and selection of one of them is then influenced by prosodic information.

To make sure that these results did not originate in some complicated, unconscious strategy by which all names for a given synonym are recalled serially and then the one most congruent with the list context is chosen, we performed a post hoc analysis of two different types of synonym pairs used in the experiments. In one type, both synonyms shared the same first morpheme, and in the other type they did not. The idea behind this was that if there are synonyms with different types of representation, and if people use an unconscious strategy to recall all possible synonym names before deciding which one to use, then the different types of representation should not have a strong effect on the extent of biasing as long as it is not harder to recall the members of one type of synonym pair than to recall those of another. Alternatively, if the biasing occurs due to feedback from prosody, then some types of synonyms may be more prone to biasing than others. In particular, we hypothesized that if the use of names for synonyms that share their first morpheme is governed by a diacritic parameter associated with a single lemma form, then these synonyms may show a different susceptibility to biasing than do those that activate two different lemmas, since the biasing would need to occur at the level of the diacritic parameter rather than at a level at which selection is made from multiple lemmas. The results showed that almost all of the biasing effect of list context came from pairs of synonyms that did not share their first morpheme. We take this as indirect evidence that in relaxed naming conditions people do not generate all possible synonym names serially and then choose the one most congruent with the list conditions.

The results can be interpreted in terms of the model of Dell et al. (1997). Because this model allows interactivity, our results might be explained by suggesting that feedback from prosodic processing could influence the activation of lemmas and, hence, the phonological word form that is likely to be chosen. If such feedback does exist, then presumably it also predicts that it should influence intervening levels (e.g., phonological word

form) between the prosody generator and the lemma level. One potential problem with this account is that the majority of words in Chinese are bisyllabic (Duanmu, 1999), although token frequencies are lower. Thus, if feedback occurs, in some cases it would be affecting the majority of words in the lexicon.

An alternative explanation of the results can be derived from the model of Caramazza (1997), according to which conceptual representations activate phonological word forms (lexemes) directly. This means that there is no need for an influence on selection through an intermediate level. Thus, for this type of model it might be easier to predict the results that we found, since word selection might be directly influenced by a prosody generator. In this case, when two different lexemes are activated, each may activate different metrical plans. List context may then influence which lexeme is articulated by influencing the speed at which the different metrical plans are activated and, hence, which one is chosen. However, we should note that how this model deals with prosody has not been discussed. In addition, the post hoc analysis might be somewhat more difficult for it to handle, although because the model has not been extended to deal with specific properties of Chinese words it is difficult to know what its exact predictions should be.

Finally, although we found an effect of prosodic biasing on lemma selection, the effect itself might be somewhat language specific. One reason for this is that it has been argued that in Mandarin Chinese prosody has a particularly strong effect in a number of different domains (Duanmu, 1999; Feng, 1998, 2002, 2003), whereas it does not appear to have such a strong effect in some other languages. One particular example of this (Feng, 2003; see also Duanmu, 1999, for other examples) is that in certain situations the grammaticality of a sentence appears to be determined by the number of syllables that a word in a specific syntactic position has (two vs. three). Although similar prosodic constraints on grammatical usage might also exist in other languages (e.g., French, in which monosyllabic adjectives occur before nouns much more frequently than bisyllabic adjectives do; see Miller, Pullum, & Zwicky, 1997, for a discussion and a possible interpretation that does not make reference to prosody), it might be that prosody constrains different aspects of different languages to varying extents. Thus, in languages such as Chinese, in which the number of syllables in a word appears to be able to govern the grammaticality of a sentence in certain circumstances, prosody might have a greater effect on the selection of synonyms with different numbers of syllables than it would in languages in which this is not the case.

REFERENCES

- BEIJING LANGUAGE INSTITUTE (1986). *Xian dai han yu pin lu ci dian* [Word frequency dictionary of Modern Chinese]. Beijing: Beijing Language Institute Press.
- 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, & Cognition*, **27**, 1430-1450.
- CUTTING, J. C., & FERREIRA, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 318-344.
- DELL, G. S., SCHWARTZ, M. F., MARTIN, N., SAFFRAN, E. M., & GAGNON, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, **104**, 801-838.
- DUANMU, S. (1999). Stress and the development of disyllabic words in Chinese. *Diachronica*, **16**, 1-35.
- FENG, S. (1998). Prosodic structure and compound words in Classical Chinese. In J. L. Packard (Ed.), *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* (pp. 197-260). Berlin: Mouton.
- FENG, S. (2002). *The prosodic syntax of Chinese* (Lincom Studies in Asian Linguistics 44). Munich: Lincom Europa.
- FENG, S. (2003). Prosodically constrained postverbal PPs in Mandarin Chinese. *Journal of Linguistics*, **41**, 1085-1122.
- FERREIRA, F. (1993). Creation of prosody during sentence production. *Psychological Review*, **100**, 233-253.
- HAYES, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
- JESCHENIAK, J. D., & SCHRIEFERS, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1256-1274.
- LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, **22**, 1-75.
- MEYER, A. S. (1994). Timing in sentence production. *Journal of Memory & Language*, **33**, 471-492.
- MILLER, P. H., PULLUM, G. K., & ZWICKY, A. M. (1997). The principle of phonology-free syntax: Four apparent counterexamples in French. *Journal of Linguistics*, **33**, 67-90.
- PACKARD, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- PETERSON, R. R., & SAVOY, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 539-557.
- RAPP, D. N., & SAMUEL, A. G. (2002). A reason to rhyme: Phonological and semantic influences on lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 564-571.
- ROELOFS, A., & MEYER, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 922-939.
- TAFT, M., & FORSTER, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning & Verbal Behavior*, **15**, 607-620.

NOTE

1. We have also performed further analyses examining the same effects, but with the proportion of monosyllabic answers given controlled for across the two groups. Very similar results were found.

APPENDIX
Chinese Names of Items in Test Pictures and Mean Percentages of
Monosyllabic Responses as a Function of Experiment and Context

Item		Percentage of Monosyllabic Responses					
Bisyllabic Name	Monosyllabic Name	Experiment 1		Experiment 2		Experiment 3	
		No Context	Context	No Context	Context	No Context	Context
雨伞	伞	67	44	69	46	80	75
宝剑	剑	86	100	100	85	100	89
铁壶	壶	44	30	83	42	63	11
电灯	灯	56	20	54	31	67	38
山羊	羊	89	70	100	69	100	100
大脑	脑	56	30	42	42	40	57
大腿	腿	100	100	85	100	100	100
轮船	船	56	60	54	33	90	57
龙虾	虾	—	—	100	44	75	71
手表	表	22	50	31	0	20	13
母鸡	鸡	100	100	92	85	100	89
手枪	枪	100	22	46	23	80	0
大炮	炮	—	—	55	30	—	—
香烟	烟	70	75	69	77	78	80
西瓜	瓜	0	0	0	9	10	0
铅笔	笔	80	56	23	15	70	22
大蒜	蒜	78	29	75	82	100	57
鸡蛋	蛋	30	33	31	33	33	38
大象	象	67	11	23	0	70	22
公路	路	70	56	62	17	88	43
汽车	车	70	22	54	0	60	11
皮鞋	鞋	80	89	92	54	70	100
天鹅	鹅	86	56	58	83	100	89
老鹰	鹰	—	—	50	78	—	—
书本	书	100	100	92	100	100	100
墙壁	墙	100	100	92	100	100	100
月亮	月	0	0	0	0	20	0
桃子	桃	—	—	67	40	—	—
眼睛	眼	11	10	15	0	0	40
耳朵	耳	0	0	0	0	0	11
链子	链	0	0	8	9	13	0
雪花	雪	13	20	30	9	25	0
鼻子	鼻	13	0	0	0	0	11
嘴唇	嘴	60	67	85	69	33	50
牙齿	牙	88	50	58	40	63	67
猴子	猴	22	22	23	8	10	33
线团	线	33	0	20	0	38	20
舌头	舌	13	13	0	0	0	0
花朵	花	100	100	92	85	100	100
鸭子	鸭	33	33	46	46	50	40
Overall mean		55	44	52	40	58	47
Mean for items not sharing first morpheme		67	50	60	45	72	53
Mean for items sharing first morpheme		39	34	39	32	37	38

Note—Dashes indicate that percentages were removed due to high error.

(Manuscript received December 17, 2003;
revision accepted for publication September 2, 2004.)