Speaking words in the second language: From semantics to phonology in 170 ms

Taomei Guo, Danling Peng

State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, PR China

Received 26 July 2006; accepted 11 November 2006
Available online 8 December 2006

Abstract

The temporal course of semantic and phonological retrieval when producing words in the second language was investigated using ERP technique. In the experiment, less proficient Chinese–English bilinguals were required to name pictures in their second language. Before performing this task, they carried out a dual choice go/nogo task on the basis of a picture’s semantic information (i.e., whether the picture was of an object or an animal) and phonological information (e.g., whether the picture’s English name starts with the letter C or F). During the time window of 200–600 ms after stimulus onset, nogo trials generated a more negative ERP profile relative to go trials. A clear-cut N200 was obtained from their difference waves. Overall, the peak latency of N200 elicited by semantic information is earlier than that by phonological information. It takes unbalanced bilinguals approximate 170 ms to retrieve phonological information after semantic information becomes available in the second language production. These findings suggest that the semantic information is encoded earlier than the phonological information during the second language production.

© 2006 Elsevier Ireland Ltd and the Japan Neuroscience Society. All rights reserved.

Keywords: Speech production; Second language; Lexical access; Temporal course; N200; ERP

1. Introduction

Language production can be simply defined as transforming concepts into sound patterns which can be produced by articulators. There is a general agreement that the core process of language production is lexical access, which mainly involves a phase that is concerned with the retrieval of semantic characteristics, and a phase that involves access to the phonological properties of the intended words (e.g., Damian, 2003; Levelt et al., 1999). An important concern in research on language production is the time when different levels of information become available.

In last several decades, studies about language production mainly focused on the mechanism of monolingual speech production. In contrast, few studies address the mechanism of lexical access in the second language production. Some existing research on second language production mostly sought help from models about monolingual speech production. So, we first reviewed theories about monolingual speech production and relevant studies.

Currently, there are two major hypotheses addressing the mechanism of lexical access of monolingual speech production. The first is the discrete two-step model (Levelt, 1999; Levelt et al., 1999), which suggests that semantic and phonological activation is independent from each other, and that a word’s semantic information is retrieved earlier than its phonological information. The alternative is the interactive activation model (Dell et al., 1999), which assumes that these two processes can spread activation to each other, and that the retrieval of phonological information may start before semantic access becomes completely available. Both these two theories got some supports from behavioral studies. For instance, Levelt et al. (1991) found that the semantic processing was earlier than phonological processing; while Damian and Martin (1999) observed that the time of phonological retrieval overlapped with the time of semantic retrieval, and that the former could start before the latter.

In recent years, ERP (event related potential) technique has also been used to examine the relationships between these two
phases involved in language production due to its high temporal resolution. Previous ERP studies (van Turennout et al., 1997; Schmitt et al., 2000; Rodriguez-Fornells et al., 2002; Rahman and Sommer, 2003; Guo et al., 2005) mainly used the dual choice go/nogo task (Miller and Hackley, 1992) to investigate the time course of semantic and phonological retrieval. In this task, subjects are instructed to classify pictures according to their semantic and phonological features by pressing a button with their left or right hand. For example, they are asked to press a left button for a picture name starting with a consonant and a right button for a picture name starting with a vowel. However, they should respond only when it is an animal but not when it is an object. Two ERP components can be obtained with this task. One is lateralized readiness potential (LRP), which reflects response preparation (e.g., van Turennout et al., 1997). The other is N200, which reflects response inhibition (e.g., Schmitt et al., 2000). The latencies of these two ERP components are suggested to reflect the relative time course of information encoding. By comparing latencies of LRP or N200 under different conditions, the temporal course of different kinds of information processing can be inferred (Schmitt et al., 2000). van Turennout et al. (1997) first used LRP to investigate the temporal course of lexical access in language production. They found that semantic information was retrieved earlier than phonological information. By using both LRP and N200, Schmitt et al. (2000) further explored the time course of semantic and phonological encoding in language production, and also observed that the retrieval of semantic features is prior to the phonological retrieval. These findings were confirmed by the study of Rodriguez-Fornells et al. (2002). In a recent study, Guo et al. (2005) extended these findings to Chinese word production, and obtained similar results. In their study, the dual choice go/nogo task was employed to examine the temporal course of the lexical access in Chinese language production. They observed that the peak latency of N200 (307 ms) elicited by semantic information was much earlier than that (447 ms) by phonological information. However, Rahman and Sommer (2003) used similar paradigm to examine similar question, they failed to obtain similar results such that these two levels of information could be retrieved in parallel.

To summarize, the studies mentioned above contribute to our understanding of the mechanism of lexical access of monolingual speech production. Compared with speech production in the native language, producing words in the second language is less automatic and has more errors (Temple, 2000). A study on slips of the tongue in the second language production revealed that some of the hypothesized mechanisms, such as backward spread activation, on monolingual speech production were not functional when speaking in the second language (Poulii, 2000). These findings have important implications that speaking in the second language probably differs from speaking in the native language in the time course of lexical access. Presumably, since bilinguals’ two languages share a common semantic representation (e.g., Kroll and Stewart, 1994), it is probably no doubt to assume that the retrieval of semantic information when speaking in either the native or the second language starts in parallel. Therefore, it is possible that the phonological encoding during producing words in the second language is less automatic and needs longer time, resulting in the fact that the semantic and phonological retrieval in the second language production tends to be two discrete steps. The present experiment is designed to investigate this possibility by using the dual choice go/nogo task. The advantage of this experimental paradigm lies in the fact that it can successfully separate the time course between different types of information processing, and that it can minimize the movement-related artifacts induced by naming aloud (e.g., Schmitt et al., 2000). Additionally, we also aim to carry out a comparison between the present study and our previous study (Guo et al., 2005) in order to explore whether speaking in the second language and the native language differs in the temporal course of semantic and phonological retrieval.

Based on extant studies (Schmitt et al., 2000; Rodriguez-Fornells et al., 2002; Guo et al., 2005), relative to go trials, nogo trials will elicit a much more negative-going ERP component at electrodes over the frontal and middle scalps. It indicates that information used for response inhibition has become available. Accordingly, the latency of N200 (the difference wave subtracted from ERPs generated by nogo and go trials) reflects the time to encode relevant information. The logic behind N200 (cf. Schmitt et al., 2000) is as follows: once information determining whether to response (go/nogo) is processed, response inhibition starts, resulting in eliciting N200. When the response hand (left or right hand to press button) is determined by semantic information, while whether to response (go/nogo) is determined by phonological information, the N200 latency reflects the time of phonological processing. When the response hand is determined by phonological information, while whether to response is determined by semantic information, the N200 latency indexes the time of semantic processing. By comparing N200 latencies of these two conditions, we can deduce the temporal course of semantic and phonological processing. Specifically, if semantic information was retrieved earlier than phonological information, then inhibition caused by semantic information would start earlier than that by phonological information. In other words, the N200 latency caused by semantic retrieval would be earlier than that by phonological processing. On the contrary, if semantic and phonological information was encoded in parallel, there would be no significant difference between the N200 latencies caused by distinct kinds of information. Furthermore, if speaking in the second language differed from speaking in the native language in the temporal course of lexical access, then there would be significant difference between the N200 latencies elicited by semantic and phonological processing of two languages. Otherwise, no any difference would be observed.

2. Materials and methods

2.1. Subjects

Fifteen right-handed undergraduate students (aged 19.4 ± 0.8 years) participated in the experiment. All subjects reported being free from neurological or psychiatric illness and had normal or corrected-to-normal vision. The study was
conducted with the written consent of each participant. They began to learn English when they were about 12 years old, and have not yet passed the CET-4 (College English Test-Band 4). This test is set to measure the English proficiency of undergraduate students in China. If one passes this test (i.e., his score is higher than 60), he is awarded a certificate indicating his scores is regular (61–84) or excellent (85–100). In order to help readers to understand how well the CET-4 corresponds to those widely known English tests such as TOEFL, we collected the CET-4 and TOEFL scores of 38 college students who have taken both two tests, and then calculated the correlation between the CET-4 and the TOEFL. We found that there was relatively high correlation between these tests ($r = 0.48$ and $p < 0.005$). We thus inferred that one might not be able to get a TOEFL score higher than 400, if he did not pass the CET-4. Additionally all the participants reported that they seldom communicated with others in English in their daily life. Therefore, it is assured that their second language is less proficient than their native language.

2.2. Materials

A set of 24 black and white line drawings of objects was selected from Snodgrass and Vanderwart (1980). According to the Chinese norms of the Snodgrass pictures (Zhang and Yang, 2003), the mean familiarity of the selected pictures in the present experiment was 4.49 (S.D. = 0.46), indicating that subjects were quite familiar with the stimuli. Half of these pictures are animals, while the remaining are objects. There are three animals starting with B, F, S and D, respectively, and three objects starting with S, C, D and F, respectively. Each picture was repeated six times. Thus, there are 144 pictures in total.

2.3. Procedure

The experiment was conducted in a soundproof room. All participants started with a picture name learning phase to ensure that they could name these pictures using these given names in the formal experiment. In the subsequent test stage, these pictures were shown on the screen one at a time, and the subjects were asked to name each picture aloud in English. They did not move on to a short practice similar to the formal experiment until they could name all of the pictures correctly. The whole learning phase took approximately 5 min.

In the informal experiment, a trial began with the presentation of a fixation cross at the center of the computer screen. The fixation cross was displayed for 1500 ms, and then substituted by a picture after a delay of 400 ms. Each picture disappeared from the screen after 500 ms. As soon as possible after the picture appeared on the screen, participants were required to perform a dual choice go/nogo task. After an interval of 1600 ms, a cue for picture naming was presented. When they saw it, they were instructed to name the picture in English in a low voice. During the experiment, participants were required not to move their bodies, and blink only at the time when the fixation cross appeared. The formal experiment lasted for approximately 1.5 h, and was divided into two parts. In each part, participants received four sets of instructions one by one. In one part of the formal experiment, the respond hand was determined by semantic information (i.e., whether it is an animal or an object), while the go/nogo execution was put into on the basis of phonological information (e.g., whether the initial phoneme of the English name of a picture is B or S). In the other part, the respond hand was determined by phonological information, while the go/nogo decision was made according to semantic information. All the eight specific instructions were shown in Appendix B. The sequence of two parts was counterbalanced across subjects. Participants could rest after they accomplished one type of instruction.

2.4. EEG recording and analysis

The continuous electroencephalogram (EEG) was recorded from 64 sintered tin electrodes mounted in an elastic cap (Quick-Cap, Neuroscan Inc.). Linked-mastoids were used as the reference for EEG electrodes. Bipolar horizontal and vertical electrooculographic (EOG) activity was recorded for artifact rejection purposes. Electrode impedances were kept below 5 kΩ. Electrical signals were amplified with Neuroscan Synamps (50 Hz notch filter), using a band-pass filter of DC–100 Hz and a sample frequency of 500 Hz. Average ERPs, from −100 to 800 ms after the presentation of pictures, were computed for trials associated with a correct response as a function of experimental condition. Trials contaminated by excessive eye-blink or movement artifacts, peak-to-peak deflections over 140 μV, were rejected, first by an automatic procedure and subsequently on visual inspection by a manual procedure prior to averaging. Averages were aligned to a −100 ms prestimulus baseline and further subjected to digital filtering with the high-cutoff set at 20 Hz.

First, the ERPs elicited by go trials were subtracted from those by nogo trials to obtain the N200 component. On the basis of visual inspection of the waveforms as well as previous literature, the distribution of N200 mainly located on the frontal and central scalp. Therefore, 11 representative electrodes (AF3, AF4, PZF, F1, Fz, F2, FC1, FC2, C1, Cz, C2) in these regions were selected for statistical analysis. Then, in the time window of 200–700 ms after onset of pictures, the peak latency of N200 was computed for each condition and for each subject. Peak latency was defined by the latency at which the maximal amplitude value occurred within the selected epoch. A three-way repeated measures ANOVAs were carried out for the lateral electrodes, and a two-way for the central electrodes. The three-way ANOVA design included one condition factor, namely the type of information (semantics and phonology), one topographical factor, namely two hemispheres, and one electrode factor. The two-way ANOVA design included one condition factor and one electrode factor. For all analyses, degrees of freedom were adjusted according to the method of Greenhouse–Geisser when appropriate. The adjusted degrees of freedom and $p$ values would be reported.

3. Results

3.1. Behavioral data

Reaction times beyond 1500 ms were excluded. Table 1 presents average response latencies and error rates for two conditions. These data were submitted to a pair-wise $t$-test. There was no significant difference in reaction times between two conditions, $t(14) < 1$, and there was also no significant effect of error rates, $t(14) = 1.8, p > 0.05$, reflecting that similar difficulty was involved in two kinds of tasks.

Furthermore, a two-way ANOVA was performed to the behavioral data of the present study and our previous study (Guo et al., 2005). One factor referred to the type of language (i.e., English or Chinese), while the other referred to the condition (i.e., semantics or phonology). The output on reaction times showed that the main effect of language was significant, $F(1, 14) = 25.13, p < 0.001$, indicating that participants made slower responses in the second language (918 ms) than in the native language (823 ms), and that the second language is less proficient than the native language; however, other effects were not significant, $F < 1$. The output on error rates revealed that, the main effect of the condition was not significant, $F(1, 14) = 3.28, p > 0.09$; the main effect of language and the interaction were also not significant, $F < 1$.

<table>
<thead>
<tr>
<th>Task</th>
<th>RT (ms)</th>
<th>ER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go/nogo = semantics</td>
<td>903 (123.57)</td>
<td>3.4 (2.51)</td>
</tr>
<tr>
<td>Go/nogo = phonology</td>
<td>933 (117.77)</td>
<td>2.3 (2.21)</td>
</tr>
</tbody>
</table>

Note: Standard deviations were illustrated in parentheses.
3.2. Electrophysiological data

Fig. 1 shows the grand average waveforms for two conditions. During the time window of 200–600 ms after stimulus onset, nogo trials generated a more negative ERP profile relative to go trials. A clear-cut N200 was obtained from their difference waves. Overall, the average of peak latencies of N200 elicited by semantic information (310 ms) over the 11 electrodes is earlier than that by phonological information (484 ms). This was confirmed by further analyses.

According to the output of the three-way ANOVA, the main effect of the condition was significant, $F(1, 14) = 46.95$, $p < 0.001$, with the N200 peak latency for semantic information earlier than that for phonological information. However, all the other effects were not significant, $p > 0.1$.

Similar pattern was observed from the output of the two-way ANOVA. Specifically, the main effect of condition was also significant, $F(1, 14) = 40.58$, $p < 0.001$, with the N200 peak latency for semantic information earlier than that for phonological information. But all the other effects did not reach any significant level, $p > 0.1$.

In order to examine whether there was any difference between producing the second language and the native language, comparisons were also carried out to the N200 data of the present study and those of Guo et al. (2005). The results were summarized as follows.

When the go/nogo decisions were made based on the semantic information, both the analyses for the lateral electrodes and the central electrodes showed that, the main effect of language was not significant, $F < 1$, indicating there is no significant difference between the N200 peak latencies elicited by semantic information of two languages. Other effects were also not significant, $p > 0.2$.

Similar pattern was obtained when the go/nogo decisions were made based on the phonological information. Specifically, both the analyses for the lateral electrodes and the central electrodes showed that, the main effect of language was not significant, $F < 1$, reflecting there is also no difference between the N200 peak latencies elicited by phonological information of two languages. Other effects were also not significant, $p > 0.3$.

![Fig. 1. The grand average ERP waveforms for go/nogo trials in the hand = semantic condition (left column), and the hand = phonology condition (middle column). The "nogo minus go" difference waves for two conditions are shown in the right column.](image-url)
4. Discussion

The dual choice go/nogo task was employed to examine the time course of lexical access in the second language production. In the experiment, subjects were asked to name pictures in their second language, but before doing this they have to perform a dual choice go/nogo task based on pictures’ semantic and phonological information. As we have noted in the introduction, one of the ERP components can be obtained in this task is N200, whose latency indicates the encoding time of related information. By comparing N200 latencies between different conditions, the temporal course of distinct cognitive processes can be inferred (Schmitt et al., 2000). The results of this study are very clear-cut. Although the subjects were trained with the picture names in English before the formal experiment and each picture was repeated six times in the formal experiment, we found that there was still significant difference between the semantic retrieval and the phonological retrieval, that is, the N200 latency elicited by semantic information was earlier than that by phonological information.

This result is consistent with most ERP studies on speech production in the native language (van Turennout et al., 1997; Schmitt et al., 2000; Rodriguez-Fornells et al., 2002; Guo et al., 2005). However, Rahman and Sommer (2003) using the same task found that, during the German language production, the semantic information and phonological information could be retrieved simultaneously. One possibility is that, the second language of the participants is less proficient than the native language, and thus needs more time to retrieve phonological information. From this perspective, as for the less proficient bilinguals, the semantic retrieval and phonological retrieval in the second language production might be independent from each other. As bilinguals become more proficient, it may take a shorter time to encode phonological information, and these two stages in producing the second language may start in parallel. Another possible reason was that subjects in the experiment of Rahman and Sommer (2003) were required to judge whether the animal in a picture was a carnivore or an herbivore, which is more difficult compared with the task to judge whether it is an animal or object. From the perspective of semantic representation, an animal (e.g., sheep) contains many semantic characteristics. Some of them (e.g., hair) are easily encoded, while others (e.g., food habit) probably need longer time to retrieve. In such case, during the language production, the easily retrieved semantic characteristics may be encoded earlier than the phonological information, whereas those difficult to be retrieved may be processed in parallel to phonological information. When considering the overall process of language production, the semantic information should be retrieved earlier than the phonological information. Our result is in line with most of the previous studies. We, therefore, suggest that the semantic information is also encoded earlier than the phonological information during the second language production.

As we have mentioned in the introduction, both of the two theories on language production claim that semantic information is processed earlier than phonological information. However, the discrete two-step theory (Levelt, 1999; Levelt et al., 1999) indicates that the semantic and phonological processes are two different stages, but the latter always starts after the former. However, the interactive activation model (Dell et al., 1999) argues that the phonological processing can start before the semantic processing is over, and these two processes are interactive. The result in the present experiment showed that semantic information was processed earlier than phonological information in the second language production, which supports the common viewpoint of two theories.

More interestingly, by comparing behavioral data of the present study and those of Guo et al. (2005), selecting words in the second language was found much slower than selecting words in the native language. Since response times include not only the time for semantic processing but also the time for phonological processing, we can not yet tell which stage in the second language production needs longer time. As mentioned in the introduction, we assume that the retrieval of semantic information when speaking either in the second language or in the native language starts in parallel because it is generally agreed that bilinguals’ two languages share a common semantic representation (e.g., Kroll and Stewart, 1994). This hypothesis was confirmed by our comparison of the N200 peak latencies elicited by semantics in two studies, which showed that there was no significant difference between the semantic retrieval of two languages. Therefore, difference in response times between the present study and our previous study might indicate that the phonological encoding when producing words in the second language is less automatic and need longer time when compared with speaking in the native language. This seems to be the case when we look at the ERP data of two studies. Specifically, we observed the difference in N200 latencies between semantic and phonological retrieval in the second language production was 174 ms, whereas the mean N200 latency difference between these two processes was relatively smaller, i.e., 140 ms, in the study on Chinese word production by Guo et al. (2005). However, out of our expectation, further comparison of the N200 peak latencies elicited by phonological information revealed that there was no significant difference between the phonological processing of two languages. One possible reason is that repetition of the limited number of pictures decreased the difference between the semantic and phonological retrieval in the second language production. A new study has already been designed to examine this probability, and we hope to obtain an evident answer in the near future.

In conclusion, the findings obtained in the present experiment indicate that semantic information becomes earlier than phonological information when producing words in the second language.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (30600179) to Taomei Guo, and the National PANDENG grant (95-Special-09) to Danling Peng. We thank the two anonymous reviewers for their valuable
comments. We also thank Chunming Lu and Hongyan Liu for their help with collecting data, and prof. Zhenhai Qi for his proofreading the manuscript.

Appendix A. List of names of the pictures used in the present experiment

<table>
<thead>
<tr>
<th>Initial phoneme</th>
<th>Animal</th>
<th>Initial phoneme</th>
<th>Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Bear</td>
<td>C</td>
<td>Cap</td>
</tr>
<tr>
<td>B</td>
<td>Bee</td>
<td>C</td>
<td>Chain</td>
</tr>
<tr>
<td>B</td>
<td>Bird</td>
<td>C</td>
<td>Comb</td>
</tr>
<tr>
<td>D</td>
<td>Deer</td>
<td>D</td>
<td>Desk</td>
</tr>
<tr>
<td>D</td>
<td>Duck</td>
<td>D</td>
<td>Door</td>
</tr>
<tr>
<td>D</td>
<td>Dog</td>
<td>D</td>
<td>Drum</td>
</tr>
<tr>
<td>F</td>
<td>Fish</td>
<td>F</td>
<td>Flower</td>
</tr>
<tr>
<td>F</td>
<td>Fox</td>
<td>F</td>
<td>Flag</td>
</tr>
<tr>
<td>F</td>
<td>Frog</td>
<td>F</td>
<td>Fork</td>
</tr>
<tr>
<td>S</td>
<td>Sheep</td>
<td>S</td>
<td>Scissors</td>
</tr>
<tr>
<td>S</td>
<td>Snake</td>
<td>S</td>
<td>Shoes</td>
</tr>
<tr>
<td>S</td>
<td>Spider</td>
<td>S</td>
<td>Stool</td>
</tr>
</tbody>
</table>

Appendix B. Eight instructions for the formal experiment

(1) Press the left button for an animal, press the right one for an object; press only when the picture name in English starts with B, but not with S.

(2) Press the left button for an animal, press the right one for an object; press only when the picture name in English starts with F, but not with C.

(3) Press the left button for an animal, press the right one for an object; press only when the picture name in English starts with S, but not with D.

(4) Press the left button for an animal, press the right one for an object; press only when the picture name in English starts with D, but not with F.

(5) Press the left button when the picture name in English starts with B, press the right one when the picture name starts with S; press only when it is an animal, but not when it is an object.

(6) Press the left button when the picture name in English starts with F, press the right one when the picture name starts with C; press only when it is an animal, but not when it is an object.

(7) Press the left button when the picture name in English starts with S, press the right one when the picture name starts with D; press only when it is an animal, but not when it is an object.

(8) Press the left button when the picture name in English starts with D, press the right one when the picture name starts with F; press only when it is an animal, but not when it is an object.

References


