BILINGUAL SWITCHING: THE PHONOLOGICAL LEVEL

A. CARAMAZZA† AND G. YENI-KOMSHIAN
Johns Hopkins University

E.B. ZURIF
Aphasia Research Center, Boston University School of Medicine, and Boston Veterans Administration Hospital

ABSTRACT

A study was conducted to discover whether Macnamara's (1967) two-switch model of bilingual functioning held at the phonological level. A group of Canadian French-English bilinguals were tested for their perception and production of the phonological feature Voice Onset Time. The subjects were tested twice, once in a French language set and once in an English language set. The perceptual functions obtained under the two language set conditions were virtually identical, whereas the production distributions were significantly different in the two conditions. The results suggest that the two-switch model can be applied to the phonological level.

Bilingual speakers appear to have no difficulty in keeping their two languages functionally distinct, which suggests that they process each language independently from the other. In fact, this notion of language-selective encoding and decoding has been formalized in Macnamara's (1967) "switching" model of bilingual functioning. The model has two switches: an output switch controlled directly by the speaker and an automatic input switch controlled by the input stimulus. The switches are thought of as simple decision mechanisms that select the language to be processed by being in one of two possible states in each language, on or off. When the switches for one language are on, those for the other are off.

The output switch in this model is relatively easy to conceptualize; a bilingual decides to speak in one language (L₁) rather than the other (L₂), and in the metaphor of the model, switches on L₁ while automatically switching off L₂. Empirical support for positing such a speaker-controlled mechanism rests with the assumption that it takes a measurable amount

*This research was supported in part by the National Research Council of Canada, Grant Number AO 291; NINDS Research Grant number 09994 and NINDS Research Grant number 06209. Part of this research was presented at the 84th meeting of the Acoustical Society of America, Miami, November 1972. We would like to thank Dr Bert Green and Mr Frank Sonsini for statistical advice; Dr Howard Egeth and Mr Jack Yates for helpful comments on an earlier version of this paper; and Drs A.S. Abramson and L. Lisker for the loan of their synthesized continua. The assistance of Mrs Rignell and Mr Carbone in data collection is appreciated.

†Requests for reprints should be sent to A. Caramazza, Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.

310

of time for one switch to be turned on in place of the other, and on this basis, the evidence is reasonably strong. Thus, bilinguals have been found to utter strings in either L₁ or L₂ at significantly faster rates than when alternating between L₁ and L₂ (Kolers, 1966; Macnamara, 1967b). However, when the subjects are given enough time to anticipate a linguistic switch, and presumably sufficient time to set their output device, then production rates for the mixed and unilingual language lists are no longer significantly different (Macnamara, Krauthammer, & Bolgar, 1968).

The input device, on the other hand, is considered to be automatically set by the structure of the stimulus. By shifting control from the listener to the stimulus, the model presupposes a preattentive mechanism (Neisser, 1967). According to the model, this mechanism generates a preliminary categorization of the signal as being either L₁ or L₂. Only when the stimulus has been categorized does the appropriate channel get switched on for subsequent processing (Macnamara, 1967a).

There is, of course, the problem of specifying the level at which there is sufficient information for perceptual language categorization. Given that there are multiple levels at which a stimulus may be analysed – phonological, syntactic, and semantic – at which level of processing does the bilingual subject decide that a stimulus belongs to one or the other of his two languages?

There is also another problem: although the theoretical characteristics of the input switch seem satisfactory, both intuitively and when applied in an ad hoc manner (Preston, 1965; Macnamara, 1967a), specific predictions based on the model have not been borne out. Macnamara and Kushnir (1971, Exp. iii) presented bilingual subjects with written sentences that alternated either randomly or predictably between languages. They reasoned that if the input switch was not automatic, subjects would process the material faster in the predictable condition where they could anticipate the language change and control the switch. However, if it was automatic, subjects would find it no more difficult to process the randomly alternating series than the predictable series. Yet, when their bilingual subjects were asked to judge the truth or falsity of these visually presented sentences, neither outcome was observed. Rather, their subjects were significantly faster in the random condition.

This rather uninterpretable result suggests that the paradigm itself be examined. It is not obvious, for example, what the effects of anticipation should be on an automatic input switch. In addition, and perhaps more importantly, the choice of visual material may not have been too appropriate. The different levels of language are not nearly as accessible in a visual stimulus as they are in an auditory one. Specifically, the unambiguous categorization of a visual input into one of two languages may only be
possible at the semantic level, whereas with an auditory stimulus there may be an immediate phonemic categorization.

Given these reservations and the possibility of phonemic language categorization, we felt that a more direct way to evaluate the two-switch model of bilingual functioning would be to do so at the phonological level. What was done experimentally, then, was to investigate how Canadian French-English bilinguals perceived and produced speech under two different conditions: when they were set to speak and be spoken to in one of their languages, and then again when they were in a set for their other language. In both conditions, perception was assessed by having the subjects label synthetic speech sounds that varied along the continuum of voice onset time (vor). For the production part of the experiment, the subjects read aloud lists of English and French words which were then spectrographically analysed for their vor values.

vor stands for the temporal relation existing between changes in the glottal aperture and supraglottal gestures. It is one of the acoustic dimensions of a speech sound and, as such, has been shown to be an important parameter for distinguishing between the voiced and voiceless forms of the stop consonants /b-p/, /d-t/, and /g-k/ (Liberman, Delattre, & Cooper, 1958; Liberman, Harris, Kimmey, & Lane, 1961; Lisker & Abramson, 1964; Lisker & Abramson, 1967). In the production of English utterances, for example, a short voicing lag is observed for the voiced phonemes and a relatively long lag for the unvoiced phonemes (Lisker & Abramson, 1964). Moreover, vor has a complementary role in speech perception; in fact, it serves as a sufficient cue for phonemic categorization in several languages (Abramson & Lisker, 1970).

For the purpose of this experiment, however, it should be emphasized that the distinctions made along the voicing dimension are not the same in Canadian English and Canadian French. In essence, vor appears to be a sufficient cue for voicing distinctions in the former language but not in the latter. Identification functions for unilingual speakers of Canadian French show both a wider range of perceptual uncertainty and a different division of the vor continuum than do those of Canadian English speakers (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973). In addition, there are production differences: English speakers show clearly non-overlapping vor distributions when uttering phonemic contrasts, whereas French speakers do not; and the voicing lags for /p, t, and k/ are significantly longer in English than in French (Caramazza et al., 1973).

It is in the context of these differences that the two-switch model of bilingual functioning permits a number of predictions. Consider first the input switch. If it is automatic, then Canadian French-English bilinguals should generate perceptual functions that are very much alike in the En-
English and French language sets. That is, given that vor can be effectively utilized as a phonemic cue only in English, only the English input switch should be activated by the voicing variants, and this should occur whether the bilinguals are prepared to hear English or French speech. On the other hand, if the input switch is listener controlled, then the perceptual functions in the two sets should be noticeably different, both in terms of monotonicity and the vor value of the voiced-voiceless crossover. That is, one would expect monotonic crossovers when the listener is prepared to hear English speech, and a wide range of perceptual uncertainty when the listener is set for French.

Finally, with respect to output, since the switch in this case is assumed to be controlled by the speaker, the bilingual subjects should be expected to produce different vor distributions in each language condition.

**Method**

**Subjects**
The subjects were 20 paid Canadian French-English bilingual high school seniors. They were all native French speakers and had begun to acquire English at no later than their seventh birthday. The selection of these subjects was arrived at on the basis of both self-report ratings of proficiency in English and reading speed in English, the latter being an especially good measure of bilingualism (Macnamara, 1969).

**Materials**
The experimental stimuli in the perception part of the experiment were made from three different sets of stop + vowel syllables synthesized at the Haskins Laboratories. For each set there were 37 syllables that varied only on vor. Thus, one set was subjectively perceived as ranging from /ba/ to /pa/; the second, as ranging from /da/ to /ta/; and the third, from /ga/ to /ka/. The duration of each syllable was 350 msec. Five random sequences of each of these three sets, or continua, were made by splicing, and these 15 sequences were used in the perception test.

The production stimuli were common English and French words separately typed on 3 x 5 white cards, and each word contained one of the stop consonants in an initial position. There were three English words and three French words for each of the six stop consonants.

**Procedure**
Subjects were tested individually in acoustically quiet rooms. Each subject participated in two testing sessions held two to three weeks apart. We attempted to create a psychological set for English in one session and, in the other, a set for French. Thus, one session was held at an English-speaking university; all extraneous conversation as well as the instructions were in English; and the experimenter was a unilingual native Canadian English speaker. In contrast, the other session was conducted in French by a native Canadian French speaker at a French high school. Half of the group was tested first at the university; the other half was tested first at the high school.

Each testing session began by having the subject read aloud a set of words that were unambiguously either French or English and that contained either member of a con-
trasting stop consonant pair. These responses were recorded on a Sony stereophonic tape recorder for later spectrographic analysis. The subject was then asked to label the vor variants for the same class of stops that he had just read. The labelling procedure consisted of having the subjects mark on a printed form their choice of either the voiced or voiceless consonant of each pair. The stimuli for this part of the experiment were delivered through binaural headphones, and consisted of the five different random orders of the continuum. This procedure of assessing production first and then perception was repeated in a random order for each of the three classes of stop consonants in each language condition.

Results

Input Switch: Perception

Figure 1 presents the identification functions for the three phoneme pairs in the two language-set conditions. In each case the percentage of “voiceless” responses, /p, t, and k/, is plotted as a function of the vor continuum, and in each case the curves generated in the two language sets appear to be clearly alike.

In order to describe the similarity of these functions statistically, voiced-voiceless crossover values were computed for each subject in both set conditions – French and English. (The mean crossover points for each of the contrasting phoneme pairs in the two language sets are presented in Table 1.) A two-factor, language set and phoneme type, repeated-measures analysis of variance was performed on these data. There were two levels in the language set factor (French and English) and three levels in the phoneme type factor (/b,g/, /d,t/, and /g,k/). The analysis revealed a highly significant main effect for phoneme type, $F(2,38) = 31.07, p < 0.001$, but no main effect for language set, $F(1,19) = 0.18, p > 0.05$, and no interaction of language set $\times$ phoneme type, $F(2,38) = 0.57, p > 0.05$. Thus, there is no evidence of any significant perceptual difference between the two language sets.

Output Switch: Production

Wideband spectrographic analyses were performed on the recordings of the spoken English and French words. The resultant vor distributions for the voiced phonemes were markedly discontinuous in both French and English language sets and so could not be analysed by parametric statistical techniques. However, all the information necessary for our test of the switching model was obtained by analysing the distributions of the voiceless phonemes. The mean vor values for these phonemes are presented in Table II.

A repeated measures analysis of variance was carried out on the voiceless phoneme distribution. The main effects of language set and phoneme type were both highly significant ($p < 0.001$). Specifically, the voicing lags were
VOICE ONSET TIME (msec)

Figure 1. Percentage of /p/, /t/ and /k/ responses as a function of VOT. The data points are based on 100 observations each.
TABLE I
CROSS OVER MEANS (msec) FOR EACH PHONEME TYPE IN THE TWO LANGUAGE SETS, FRENCH (BF) AND ENGLISH (BE)

<table>
<thead>
<tr>
<th></th>
<th>/b-p/</th>
<th>/d-t/</th>
<th>/g-k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>17</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>BE</td>
<td>16</td>
<td>23</td>
<td>29</td>
</tr>
</tbody>
</table>

TABLE II
MEAN VOT VALUES (msec) FOR THE VOICELESS CONSONANTS IN THE TWO LANGUAGE SETS FOR THE PRODUCTION TASK

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>20</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>BE</td>
<td>39</td>
<td>48</td>
<td>67</td>
</tr>
</tbody>
</table>

much longer for words spoken in English than for those spoken in French and, as expected, voicing lags for phoneme types were ordered as /k/>/t/>/p/. The interaction of language set and phoneme type was also significant, reflecting a wider crosslanguage difference for the phoneme /k/ relative to the other two voiceless phonemes.

DISCUSSION

The present experiment attempted to clarify the nature of language switching in bilingualism. Two points emerged from the data: first, that language switching is under different sources of control at the input and output stages; and second, that language categorization can occur at the phonological level.

Consider first the perception data. Given that no effective language-specific separation was obtained, it would appear that the input switch in the Macnamara model is stimulus controlled. That is, even though vor does not appear to be a phonemic cue in Canadian French, the bilingual subjects were not necessarily restricted to judgments based on an English phoneme analyser. After all, unilingual Canadian French speakers do generate consistent identification functions when labelling vor variants (Caramazza et al., 1973); it is just that their strikingly non-monotonic functions suggest a wide range of perceptual uncertainty. So, if the input switch had been controlled by the listener, the bilingual subjects should have generated very different perceptual functions in the French and English conditions.
The production results, on the other hand, support the view that the language output device is speaker-controlled; also in accord with Macnamara’s two-switch model. That is, as predicted by the model, control of the language output device seems to lie entirely with the speaker as evinced by the consistently longer VOR values produced in the English set relative to the French set. Yet, although bilinguals seem able to effect a phonological separation in their speech, they are not totally alike unilingual speakers of each language. As we have pointed out elsewhere (Caramazza et al., 1973), French-English bilinguals, while markedly more sensitive to VOR than unilingual Canadian French speakers, are still less sensitive to this cue than those who speak only English. Bilinguals, then, may acquire two independent linguistic systems, but each seems to modify the other during the course of development.

Finally, and of special interest, is our finding that the control of each switch extends to the phonological level. This would appear to be particularly important at the input stage. If preattentive mechanisms need not rely on semantic information to categorize the language being spoken, the system that emerges is a highly efficient one whereby the input can be language—tagged very early in the processing stage.

RéSUMÉ

Etude sur la valeur du modèle de bilinguisme à double aiguillage de Macnamara (1967), appliqué au niveau de la phonologie. Des sujets canadiens bilingues (anglais-français) subissent un test de perception et de production phonologique suivant la technique Voice Onset Time. Les sujets subissent le test deux fois, une fois dans un cadre de langue française et une fois dans un cadre de langue anglaise. Les fonctions perceptives observées dans ces deux conditions sont virtuellement identiques, alors que les distributions des activités de production diffèrent significativement d’une condition à l’autre. Ces résultats suggèrent que le modèle à double aiguillage peut s’appliquer au niveau de la phonologie.

REFERENCES


(First received 27 September 1973)