LANGUAGE AND COGNITIVE PROCESSES, 2003, *18* (2), 175–204

Shared phonological encoding processes and representations of languages in bilingual speakers

Ardi Roelofs
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Four form-preparation experiments investigated whether aspects of phonological encoding processes and representations are shared between languages in bilingual speakers. The participants were Dutch–English bilinguals. Experiment 1 showed that the basic rightward incrementality revealed in studies for the first language is also observed for second-language words. In Experiments 2 and 3, speakers were given words to produce that did or did not share onset segments, and that came or did not come from different languages. It was found that when onsets were shared among the response words, those onsets were prepared, even when the words came from different languages. Experiment 4 showed that preparation requires prior knowledge of the segments and that knowledge about their phonological features yields no effect. These results suggest that both first- and second-language words are phonologically planned through the same serial order mechanism and that the representations of segments common to the languages are shared.

INTRODUCTION

Bilingual speakers are persons who regularly use two or more languages for their verbal communication. Many aspects of bilingualism have been studied, ranging from the acquisition of a second language (Klein, 1986) to the representation of the two languages in the brain (Albert & Obler, 1978; Kim, Relkin, Lee, & Hirsch, 1997) and bilingual aphasia (Paradis, 1995).

Requests for reprints should be addressed to Ardi Roelofs, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, The Netherlands.
Email: ardi@mpi.nl

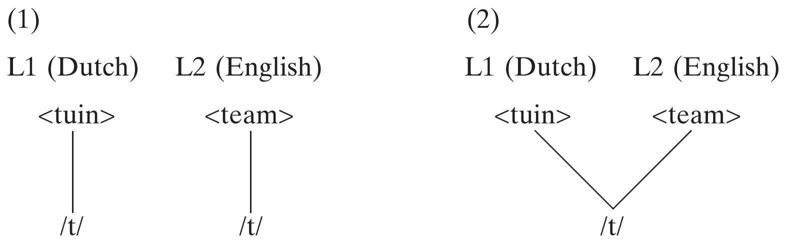
I am indebted to Daniëlle van den Brink for her help in preparing and running the experiments, and to Antje Meyer, Pim Levelt, Daan Hermans, Tamar Gollan, and two anonymous reviewers for their helpful suggestions.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000515

Most bilinguals use their languages for different purposes and in different situations, and therefore balanced bilingual speakers, those who speak both languages equally fluent in all contexts, are probably the exception (Grosjean, 1982; Leopold, 1939; Mackey, 1967; Weinreich, 1953).

A question that has received increasing attention in recent years concerns the mental lexicon of bilingual speakers (De Groot & Kroll, 1997; Schreuder & Weltens, 1993). In planning speech, speakers draw on the knowledge about words that they have stored in long-term memory. This stored information comprises the meaning of the words, their syntactic properties, and information about their form including the morphemes and segments. The question is at which levels the two languages share processes and representations, where they are merely linked to each other, and where they are separate. Most of the research concerning this issue has concentrated on the representation of word meaning (for early studies, Caramazza & Brones, 1979; Ervin & Osgood, 1954; Weinreich, 1953; for reviews of recent studies, De Groot, 1993; Francis, 1999; Kroll, 1993). The question of whether there are shared processes and representations at the level of word forms, however, has been neglected. Thus, ironically, one of the most immediately obvious aspects of bilingual speech has been least intensively researched—Watson (1991) called phonology "the Cinderella of bilingual studies" (p. 25). Studies typically looked at the perception (Altenberg & Cairns, 1983; Grainger, 1993; Grainger & Dijkstra, 1992; Jared & Kroll, 2001) or the articulation of words in the first and second language (Flege, 1991; Flege & Eefting, 1987), but the access of word-form information in memory for production has not received much attention. The present paper addresses this issue.

In phonological encoding, a speaker recovers the phonemic segments of morphemes from memory and uses the segments to assemble phonological representations of the words to be spoken (Dell, 1986; Levelt, 1989, 1992; Roelofs & Meyer, 1998). Morphemes differ between languages, but some phonemic segments are common (Kenstowicz, 1994; Ladefoged, 2001; Ladefoged & Maddieson, 1996) and hence their representations may, in principle, be shared between languages in memory. It has been estimated that there are about 7000 languages in the world, but only about 200 different vowels and some 600 different consonants (Ladefoged, 2001; Ladefoged & Maddieson, 1996). This means that languages must have equivalent vowels and consonants (this is also the idea behind the international phonetic alphabet, IPA; Pullum & Ladusaw, 1996). For example, whereas the English language has the segment /θ/ (as in *thumb*) and the Dutch language does not (Booij, 1995), both languages have segments such as /t/ (i.e., a voiceless alveolar plosive that is contrastive within the language). Segments such as /t/, /p/, and /k/ occur with only slight differences in about 98% of the world's languages (Ladefoged, 2001). This raises the question whether the memory representation of common segments is also shared between languages or whether a bilingual speaker has different memory tokens of a segment for each language. As concerns the relation in long-term memory between language-specific morphemes (e.g., Dutch <tuin>, *garden*, and English <team>) and segments common to both languages (e.g., /t/), there are therefore at least two theoretical positions (where L1 stands for the speakers' first language and L2 for their second):



Weinreich (1953) referred to positions (1) and (2) as "co-ordinate" and "compound", respectively. He believed that the distinction could be relevant for many levels of language such as phonology, syntax, and semantics, although later applications by researchers of these terms have been restricted to separate or shared representations of meaning. Position (1) holds that the English and Dutch morphemes point to separate token representations of a common segment type in memory, whereas position (2) maintains that the morphemes point to the same token representation. Recently, MacSwan (2000) argued on theoretical grounds that due to the highly language-specific character and the nature of the rule system, the phonological systems of a bilingual speaker cannot be shared. Empirical evidence that could possibly be relevant for deciding between the theoretical positions (1) and (2) comes from speech error analyses, phonetic studies, and neuroimaging. The evidence is inconclusive, however.

In speech errors, languages may get interwoven in very complicated ways. Sometimes bilingual speakers erroneously use segments or apply phonological rules of their native language when speaking a foreign language. In producing English, for instance, Dutch speakers sometimes use English segments in accidentally accessed Dutch words (see Poulisse, 1990, 1999, for an extensive discussion of bilingual speech errors). An example is the Dutch word *stuk* (/stuk/, where the vowel has the feature [−back]) meaning *piece* pronounced as the English word *stuck* (/stuk/, where the vowel has the feature [+back]). Such interference suggests that the phonological systems of the first and second language are not entirely separate. It is unclear, however, whether the error is due to shared representations or to closely linked ones. Also, it is unclear at what level the error occurs. The error may be a failure in addressing language-specific features (i.e., [[−back] for Dutch and [+back] for English) from a shared abstract segment in memory, or alternatively, if the segment representation is not shared, a failure in selecting the right abstract segment. To conclude, the speech error evidence is inconclusive about whether there are shared or separate phonological processes and representations.

Whereas the phonological encoding level has not been a domain of much empirical investigation, several studies have looked at the phonetic level, in particular, the pronunciation of segments in the different languages. The results of several studies suggest that certain aspects of languages may be shared between the different languages in the minds of bilinguals. In particular, it appears that few late learners of a second language fully differentiate their pronunciation of common segments in their two languages (Flege & Eefting, 1987; Flege & Port, 1981). For example, Flege (1991) observed that native Spanish speakers who acquired English as a second language as adults produced intermediate values for the voice-onset time (VOT) of an English /t/ and a Spanish /t/. However, Spanish speakers who acquired English early, by the age of 5 or 6 years, produced values identical to monolingual speakers of the two languages. Flege proposed that adult learners implement the phonetic category of their second language by applying a specific phonetic realisation rule to the phonetic category of their first language, whereas early learners establish different phonetic categories for each language. However, the issues at the phonetic level where the segments are phonetically realised are independent of those at the phonological level, inasmuch as segments, either shared or not, may be mapped onto shared or different phonetic categories. Segments at the phonological level represent the contrastive sounds of a language, abstracting away from noncontrastive phonetic aspects that may vary within (e.g., coarticulation, dialect) and between languages (yielding an accent in bilinguals). Thus, phonetic studies are also inconclusive about shared or separate phonological encoding processes and representations.

Neuroimaging studies have revealed that, at least for high-proficiency bilinguals performing language comprehension tasks, the neuroanatomical representations of both languages of a bilingual are overlapping (Perani, Paulesu, Sebastian Galles, Dupoux, Dehaene, Bettinardi et al., 1998). In a functional magnetic resonance imaging study examining language production, Kim et al. (1997) had early and late bilinguals silently describe what they had done during the previous day. The brain activations for the descriptions in the first and second language were compared. Kim et al. found that in late bilinguals, the activation during second language use was spatially separated from the activation during first language use in left inferior frontal cortex (Broca's area). For early bilinguals, however, activations during first and second language use occurred in common frontal areas. In both early and late bilinguals, Wernicke's area showed little separation between languages. According to Kim et al., these results suggest that "representations of languages in Broca's area that are developed by exposure early in life are not subsequently modified. This could necessitate the utilisation of adjacent cortical areas for the second language learned as an adult" (p. 173).

However, although Broca's area is associated with phonological encoding in spoken word production (Indefrey & Levelt, 2000), the study of Kim et al. (1997) is inconclusive. Overlapping or separate activations of brain areas cannot be directly equated with overlapping or separate language representations and processes (cf. Perani et al., 1998). Furthermore, comparing discourses in the two languages, as Kim et al. did, fails to distinguish between phonological and phonetic levels, and between these and other levels in language production. Thus, it is unclear what exactly the spatially overlapping and separate brain activations reflect.

The present study taps into phonological representation and planning processing using a chronometric technique that has been shown to be able to diagnose whether processes and representations are shared among different stimuli, namely the form-preparation paradigm developed by Meyer (1990, 1991). It tests between the co-ordinate and compound views on segments mentioned earlier (positions 1 and 2 above), and examines an aspect of their planning. The remainder of the paper is organised as follows. I start by briefly reviewing how phonological encoding happens in the WEAVER++ model (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1993, 1994, 1996a, 1996b, 1997a, 1997b, 2003), which is a computational model of spoken word production (the model has been implemented for Dutch) that can provide a theoretical framework for bilingual speech production (Roelofs, 1998b, 2000, 2003). WEAVER++ shares many assumptions with other models of form encoding (Dell, 1986). However, as explained in Roelofs (1997a, 1997b), most of these other models cannot account for encoding times (the data from the present experiments), whereas WEAVER++ can. Next, I review a number of important chronometric findings concerning phonological encoding in first language production, thereby describing the form-preparation paradigm, and I discuss how WEAVER++ accounts for these findings. Then I derive, in a general fashion, a number of predictions from shared versus separate phonological encoding processes and representations, which are tested in four new preparation experiments. The paper ends with a discussion of the implications of the finding.

Phonological encoding in WEAVER++

WEAVER++ is described in detail in many other places and I refer to these publications for a discussion of the model and its empirical support (Levelt et al., 1999; Roelofs, 1992, 1997a). Word-form encoding in WEAVER++ comprises three major steps, namely morphological, phonological, and phonetic encoding. Given a morpheme such as Dutch <tuin> (*garden*), the phonological encoder recovers the corresponding segments (i.e., /t/, /ʌy/, /n/) from memory and produces a phonological representation. On the basis of this representation, the phonetic encoder generates a detailed and context-dependent form representation, which specifies the articulation tasks to be achieved. The mental lexicon is conceived of as a network of information that is accessed by spreading activation. Relevant for phonological encoding is that activation spreads from morphemes to segments. Activation of parts of the network triggers production rules that select nodes. Phonological production rules select the segment nodes linked to the morpheme nodes and syllabify the segments in a rightward fashion to construct phonological word representations. These representations specify the syllables and the stress pattern (for polysyllabic words). Phonetic production rules select syllable-based articulation programs that encode the phonological words. During this final step, the phonological features (articulatory gestures) of the syllables become available. As concerns phonological encoding, the segments and aspects of the procedures involved in constructing phonological words may be shared between languages in bilinguals, which is the issue addressed here.

WEAVER++ provides for a suspend-resume mechanism that supports rightward incremental phonological encoding. Incrementality means that encoding processes can be triggered by a fragment of their characteristic input. In phonological encoding, syllabification of a word can start as soon as the first few segments are available. The resulting partial representation can be buffered until the missing segments are available and syllabification can continue. Thus, when given partial information, computations are completed as far as possible, after which they are put on hold. When given further information, the encoding processes continue from where they stopped. Buffered forms in WEAVER++ are only expandable toward the end of the word.

The on-line form preparation paradigm

Typical tasks in the psycholinguistic study of bilingualism are reading aloud, picture naming, and translating (Snodgrass, 1993). In investigating the issue of shared processes and representations in phonological encoding, however, I have employed a preparation task, which was originally developed by Meyer (1990, 1991) for monolingual production (cf. Donders, 1868). An advantage of this task compared with the more widely used picture naming task is that the responses do not need to be names of depictable entities, which gives more freedom in the selection of materials. More importantly, as with the translation task, the form-preparation paradigm allows for switching between languages from trial to trial. What language is required for the response is cued in a natural way. The paradigm does not involve reading words aloud, so it does not involve response generation directly from visual word recognition processes, which brings it closer than reading aloud to the processes involved in spontaneous speech.

The preparation paradigm has been described in depth in various other places, and I refer to these publications for an extensive discussion and motivation of it (see especially Meyer, 1990, and Roelofs, 1998a). Meyer (1990, 1991) had developed the preparation paradigm for studying phonological encoding in a single language. In her experiments, native speakers of Dutch first learned small sets of Dutch word pairs such as *zaag–hamer*, *sneeuw–hagel*, and *arend–havik* (English *saw–hammer*, *snow–hail*, and *eagle–hawk*). During the following test phase, they had to produce the second word of a pair (e.g., *hamer*) upon visual presentation on a computer screen of the first word (*zaag*), called the prompt. On each trial, one of the prompts was presented. The order of prompts across trials was random. The production latency, the interval between prompt onset and speech onset, was the main dependent variable. Each experiment contained two types of sets, called homogeneous and heterogeneous sets. In a homogeneous set, the response words shared part of their form, for example the first syllable, as in *hamer*, *hagel*, *havik*, or the second syllable, as in *hamer*, *zomer*, *roemer*. In the heterogeneous sets, the response words were unrelated in word form. Regrouping the pairs from the homogeneous sets created the heterogeneous sets. Therefore, each word pair was tested both under the homogeneous and the heterogeneous condition, and all uncontrolled item effects were kept constant across conditions.

Meyer found shorter production latencies in homogeneous than in heterogeneous sets, henceforth called the *preparation effect* (Donders, 1868, made a similar observation using a similar task). This preparation effect was only obtained when the response words in homogeneous sets shared word-initial segments, but not when they shared word-final segments. The magnitude of the preparation effect increased with the number of shared word-initial segments. This suggests that the facilitation from homogeneity is due to preparation of word production rather than to general memory retrieval processes. Research on paired-associate learning has shown that form overlap helps memory retrieval independent of the place of overlap (see Meyer, 1990, for a review of the memory literature). Furthermore, immediate verbal recall is hampered (i.e., a lower rather than a higher level of recall is observed) when the items are similar in sound or articulatory characteristics (see Baddeley, 1997, for a review). Thus, the findings from the memory literature are opposite to the results of Meyer (1990, 1991), which rules out a general memory account of her results.

Meyer's (1990, 1991) findings have been replicated not only for several types of other morphologically simple words, but also for complex ones as well as for phrasal constructions (Roelofs, 1996a, 1997a, 1998a). Furthermore, Roelofs and Meyer (1998) showed that preparation is due to shared segments and that syllable structures and stress patterns cannot be prepared. Evidence for seriality in phonological encoding has also been obtained with other experimental paradigms such as picture-word interference (Meyer & Schriefers, 1991), where spoken primes are presented during picture naming.

According to WEAVER++, faster latencies in homogeneous sets than in heterogeneous sets follow when participants prepare and buffer partial phonological representations of the response words before prompt presentation. The confinement of the facilitatory effect to begin-related homogeneous sets (e.g., Meyer, 1990) reflects the suspend-resume mechanism that underlies the rightward incremental planning of utterances. Assume that the set of response words consists of *hamer*, *hagel*, and *havik*. Before the beginning of a trial, the phonological encoder can construct the first phonological syllable $(\text{ha})_\sigma$. As soon as a prompt (e.g., *zaag*) is given, the associated target morpheme will be retrieved (for *zaag* this is <hamer>). Segmental spell-out makes all segments of this morpheme available including those of the second syllable, and the phonological encoder can start working on the second syllable. In the heterogeneous condition (*havik*, *zomer*, etc.), nothing can be prepared before prompt presentation. There will be no advance phonological encoding. In the end-related homogeneous condition (*hamer*, *zomer*, *roemer*) nothing can be encoded in advance either. Although the segments of the second syllable are known, the corresponding part of the phonological form cannot be computed in advance because the missing segments precede the suspension point. In WEAVER++, this means that after prompt presentation syllabification must restart from the first segment of the word, which amounts to restarting the whole process. Thus, a facilitatory effect for the homogeneous relative to the heterogeneous condition is only obtained for begin-related response words. Computer simulations of Meyer's (1990) experiments can be found in Roelofs (1994, 1997a).

Overview of the experiments

The participants in the present experiments were late, relatively fluent but unbalanced Dutch–English bilinguals (i.e., having Dutch as the first, native language and English as the second language), randomly selected from the pool of participants of the Max Planck Institute. They were young adults with ages ranging from about 18 to 33 years, who had started reading, speaking, and writing English in the last year of elementary education or in the first year of secondary education. All participants began to speak English between 9 and 13 years, on average at 11.4 years. Thus, none of the participants acquired English as a child. Most likely, they picked up some English from the media before formally learning it at school for 4 to 6 years and using it in their study or job. None of the participants had acquired English without an accent. Each person only took part in one experiment.

Experiment 1 tested for rightward phonological encoding of second language words by these speakers. If a suspend-resume mechanism is shared between languages, then phonological encoding in the second language should exhibit the same seriality phenomenon that Meyer (1990, 1991) observed for the native language. Experiments 2 and 3 tested for shared representations of segments that are common to both languages (e.g., /s/ and /t/). If representations of common segments are shared between languages, speakers should be able to plan initial segments that are common to both languages without knowing the language of the word. A preparation effect should be obtained when initial segments are shared even when the languages of the words in a set differ. For example, it should be possible to plan /st/ when the set includes English *steam* and *stone* and Dutch *stoel* (*chair*).

For the native language, it has been observed that preparation effects are only obtained when the responses share their initial segments fully, but no preparation effect is obtained when the initial segments share most of their phonological features (Roelofs, 1999). For example, the production latency is smaller in sets with segment overlap (e.g., *boot*, *beer*, *boek*) than in sets with feature overlap (e.g., *boot*, *beer*, *pels,* where /b/ and /p/ share all features except voicing), whose latency does not differ from that of sets without segment or feature overlap. When one phonological feature differs, the preparation effect is completely absent. The same results were obtained not only when the critical feature was voicing but also when it was a place of articulation feature, such as whether the segments were pronounced labial or coronal (e.g., labial /v/ versus coronal /z/). Experiment 4 tested whether this also holds for the second language. Only when preparation in the second language requires segment overlap, as it does in the native language, would a preparation effect in case of mixing languages in a set (in Experiments 2 and 3) support the conclusion that the representations of common segments are shared between the languages.

EXPERIMENT 1: ALL ENGLISH MATERIALS

In this first experiment, it was examined whether phonological encoding in a second language exhibits the same seriality phenomenon that Meyer (1990, 1991) observed for phonological encoding in the native language. Are bilinguals able to plan the first syllable of a disyllabic foreign word without knowing its second syllable, but not the second syllable without also knowing the first syllable? Unbalanced bilinguals are more fluent in their first than in their second language. A difference in fluency may mean that the serial ordering mechanism works differently for the second than for the first language. Dell, Burger, and Svec (1997) provided evidence for a difference in serial ordering between highly practiced and less practiced serial behaviour. They observed that when serial behaviour becomes automatised, the types of errors that are made change. In particular, the more practiced the skill, the more anticipatory rather than perseveratory the errors. According to Dell et al. (1997), this suggests a reorganization of the underlying serial ordering mechanism. For unbalanced bilinguals, the languages have received different amounts of practice. This may imply differences between the planning of first and second language words. First, it may be that the serial ordering mechanism is still more adjustable or adaptable for second than for first language words such that non-initial fragments may be planned before initial ones in the second but not in the first language. Second, it may be that the planning of second language words is not automatised enough to allow for any advance encoding. In contrast, if the serial ordering mechanism is shared between the first and second language, such differences are not to be expected.

Method

Participants. Experiment 1 was carried out with 24 Dutch participants with ages ranging from 18 to 29 years and who started to speak English at 11.1 years, on average.

Materials and design. The materials for all experiments were obtained from the Dutch and English part of the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). All prompts and response words were nouns because suitable items were easiest to find in this word class. The materials of Experiment 1 consisted of two practice sets and twelve experimental sets of three prompt-response word pairs each (see Table 1). All response words were disyllabic. Each set was tested in a separate block

TABLE 1
Prompt-response pairs of Experiment 1

Place of overlap	*Context*	*Set*
Begin	Homogeneous	Set 1: fruit–melon, iron–metal, grass–meadow
		Set 2: comrade–fellow, river–ferry, bird–feather
		Set 3: sponge–bucket, rabbit–bunny, milk–butter
Begin	Heterogeneous	Set 4: rabbit–bunny, fruit–melon, comrade–fellow
		Set 5: river–ferry, milk–butter, iron–metal
		Set 6: grass–meadow, bird–feather, sponge–bucket
End	Homogeneous	Set 7: keys–pocket, pass–ticket, ball–racket
		Set 8: stone–pebble, desk–table, church–bible
		Set 9: soldier–fighter, mail–letter, liquid–water
End	Heterogeneous	Set 10: desk–table, keys–pocket, liquid–water
		Set 11: ball–racket, soldier–fighter, church–bible
		Set 12: mail–letter, stone–pebble, pass–ticket

of trials. In six experimental sets (the homogeneous sets) the response words shared a syllable, and in the remaining six sets (the heterogeneous sets) they were unrelated in form. Thus, in the homogeneous condition, each response word was tested together with other response words with a common syllable, whereas in the heterogeneous condition, the response words tested together in a block did not share a syllable. Following Meyer (1990), the first independent variable—homogeneous versus heterogeneous sets—will be called *context*. The same prompt-response word pairs were tested in the homogeneous and heterogeneous condition. Only their combinations into sets differed.

In three homogeneous sets, all response words shared the first syllable, and in the corresponding heterogeneous sets they did not. The shared first syllables were [me], [fe], and [b∧]. In the other three homogeneous sets, the response words shared the second syllable, and in the corresponding three heterogeneous sets they did not. The shared second syllables were [kət], [bəl], and [tər]. The second independent variable, which had two levels (begin, end), will be called *place of overlap*.

Each response word was coupled with a prompt that the author and the experimenter considered to be a strong and unambiguous retrieval cue for the corresponding target. For instance, the target *melon* was coupled with the prompt *fruit*.

Each participant was tested once on each set. Each of the three prompt-response word pairs of a set was tested eight times within a block of trials. The order of testing the word pairs was random, except that immediate repetitions of pairs were excluded. A different order was used for each block and each participant. The order of the sets was fully counterbalanced across participants. Half the participants were first tested on the sets in the begin condition, and then on those in the end condition. For the remaining participants, the order of testing the begin and end conditions was reversed. Half the participants were first tested on the homogeneous sets, then on the heterogeneous ones, and for the other half, the order of homogeneous and heterogeneous sets was reversed.

Procedure and apparatus. The participants were tested individually in all experiments. The participants were seated in a quiet room in front of a computer screen (NEC Multisync30) and a microphone (Sennheiser ME40). After the participant had read the instructions, two practice blocks (with the same structure as the experimental blocks but with different items) were administered followed by the twelve experimental blocks. In the learning phase before each block, the three word pairs of a set were presented on the screen. As soon as the participant indicated having studied the pairs sufficiently, the experimenter started the test phase. The structure of a trial was as follows. First, the participant saw a warning signal (an asterisk) for 500 ms. Next, the screen was cleared for 500 ms, followed by the display of the prompt for 1500 ms. The asterisk and prompt were presented in white on a black background. Finally, before the start of the next trial there was a blank interval of 500 ms. Thus, the total duration of a trial was 3 seconds. A Hermac computer controlled the experiment.

Analyses. The response coding and analyses were the same in all experiments. After each trial, the experimenter coded the response for errors. The experimenter judging the correctness of each pronunciation of the English and Dutch words was skilled in doing this. She was a native speaker of Dutch and learned English at school, after which she studied it at university. Experimental sessions were recorded on audio tape by a Sony DTC55 DAT recorder. The recordings contained the participants' speech and tones indicating the onset of the prompt (1 kHz) and the moment when the voice key was triggered (2.5 kHz). The experimenter heard these tones via closed headphones. The recordings were consulted after the experiment whenever the experimenter was uncertain about whether a response was fully correct. Five types of incorrect responses were distinguished: wrong response words, wrong pronunciation of the words (e.g., interference from Dutch), disfluencies (stuttering, within-utterance pauses, repairs), triggering of the voice key by non-speech sounds (noise in the environment or smacking sounds participants produced with the lips or tongue), and failures to respond within 1500 ms after prompt presentation. Incorrect responses were excluded from the statistical analysis of the production latencies. For all experiments, analyses of variance were performed on the error rates using the same design as for the production latencies. No main effect or interaction was significant both by participants and items in any of the experiments (all *p*s > .05). Therefore, I report the means for the errors but not the test statistics.

Results and discussion

Table 2 gives the mean production latencies and the error percentages for Experiment 1. The column labelled "preparation" indicates the difference between the homogeneous and heterogeneous conditions. The table shows that facilitation from segmental overlap was obtained for the begin but not for the end condition. The latencies were submitted to by-participant and by-item analyses of variance with the crossed variables context and place of overlap. Both variables were tested within participants. Context was tested within items and place of overlap between items. There was an interaction between context and place of overlap, $F_1(1,23) = 11.16$, $MSE = 835$, $p < .003$, $F_2(1,16) = 9.41$, $MSE = 371$, $p < .007$. In tests of simple effects, the effect of context was significant for the begin condition, $F_1(1,23) = 17.82$, $MSE = 797$, $p < .001$, $F_2(1,8) = 15.65$, $MSE = 341$, $p < .004$, but not for the end condition, *F*s < 1. The main effect of context was not reliable, $F_1(1, 23) = 2.88$, $MSE = 1805$, $p > .10$; $F_2(1, 16) = 5.25$, $MSE = 371$, $p < .04$, nor was there a main effect of place of overlap, $F_1(1,23) = 6.29$, $MSE = 2365$, $p < .02$, $F_2(1,16) = 2.12$, $MSE = 2626$, $p > .16$.

To summarise, in second-language production, participants could plan initial segments of a word without knowing the remainder of a word, but they could not plan non-initial segments of a word without also knowing the preceding segments. Thus, the seriality effect in form preparation first reported by Meyer (1990, 1991) for first language production was replicated for the production of second language words. The fact that the seriality effect has been obtained for both the native and the second language suggests that the first and second language are planned in the same rightward incremental fashion. Even for the relatively fluent but

TABLE 2
Mean production latencies (M, in milliseconds), error percentages (E%), and preparation effects per context and place of overlap for Experiment 1

	Context					
	Homogeneous		*Heterogeneous*		*Preparation*	
Place of overlap	*M*	*E%*	*M*	*E%*	*M*	*E%*
Begin	730	3.5	764	2.6	–34	0.9
End	774	3.1	769	3.3	5	–0.2

unbalanced bilinguals tested, there is no evidence that the serial ordering mechanism for producing second language words works differently than for producing first language words. This suggests that the mechanism is shared between languages.

EXPERIMENT 2

The second experiment tested for shared representations of segments that are common to both languages (Ladefoged, 2001; Ladefoged & Maddieson, 1996). If segment representations are shared, speakers should be able to plan initial segments that are common to both languages without knowing the language of the word. That is, a preparation effect should be obtained when initial segments are shared even when the language of the words in a set differs. For example, it should be possible to plan the onset cluster /st/ when the set includes the English words *steam* and *stone* and the Dutch word *stoel* (*chair*). By contrast, when the segment representations are not shared between languages (i.e., when an English /s/ and a Dutch /s/ are different segment tokens in memory), the variable-language sets would be segmental heterogeneous sets. Then, a preparation effect should not be obtained. It has been shown that when one word in a set begins with a different segment than the other words, the preparation effect completely disappears (Roelofs, 1999).

Method

Participants. The experiment was carried out with 32 Dutch participants with ages ranging from 18 to 30 years. They started speaking English at 11.5 years on average.

Materials and design. The materials consisted of four practice sets and twenty-four experimental sets of three prompt-response word pairs each (see Table 3). Each set was tested in a separate block of trials. In 12 experimental sets (the homogeneous sets) the response words shared the first two consonants. The shared onset clusters were /st/, /bl/, and /fl/. In the remaining 12 sets (the heterogeneous sets) no such initial segments were shared. This is the first independent variable *context* (homogeneous, heterogeneous). In six homogeneous sets and the corresponding heterogeneous ones, all prompt-response pairs were in the same language. In the remaining six homogeneous sets and the corresponding heterogeneous ones, two of the three prompt-response pairs were in one language and the third pair was in the other. This second independent variable, which had two levels (constant, variable), will be called *blocking*. In the constant-language condition, all prompt-response pairs were in English for three of the homogeneous sets and the corresponding heterogeneous ones. In the

TABLE 3
Prompt-response pairs of Experiment 2. Approximate English translations of the Dutch materials are given in parentheses

Blocking	*Language*	*Context*	*Set*
Constant	English	Homogeneous	Set 1: vapour–steam, fashion–style, pebble–stone Set 2: knife–blade, shock–blow, wound–blood Set 3: bug–fly, level–floor, pole–flag
Constant	English	Heterogeneous	Set 4: wound–blood, vapour–steam, bug–fly Set 5: pebble–stone, level–floor, knife–blade Set 6: pole–flag, shock–blow, fashion–style
Constant	Dutch	Homogeneous	Set 7: ijzer–staal, tafel–stoel, punt–stip Set 8: wang–blos, kelk–bloem, trommel–blik Set 9: muziek–fluit, kurk–fles, deel–flard (Set 7: iron–steel, table–chair, dot–point Set 8: cheek–blush, calyx–flower, box–tin Set 9: music–flute, cork–bottle, part–fragment)
Constant	Dutch	Heterogeneous	Set 10: punt–stip, muziek–fluit, kelk–bloem Set 11: deel–flard, wang–blos, ijzer–staal Set 12: trommel–blik, tafel–stoel, kurk–fles
Variable	English, Dutch	Homogeneous	Set 13: fashion–style, tafel–stoel, pebble–stone Set 14: shock–blow, trommel–blik, wound–blood Set 15: level–floor, kurk–fles, pole–flag
Variable	English, Dutch	Heterogeneous	Set 16: level–floor, trommel–blik, fashion–style Set 17: pebble–stone, kurk–fles, shock–blow Set 18: wound–blood, tafel–stoel, pole–flag
Variable	Dutch, English	Homogeneous	Set 19: punt–stip, vapour–steam, ijzer–staal Set 20: wang–blos, knife–blade, kelk–bloem Set 21: deel–flard, bug–fly, muziek–fluit
Variable	Dutch, English	Heterogeneous	Set 22: muziek–fluit, knife–blade, punt–stip Set 23: ijzer–staal, bug–fly, wang–blos Set 24: kelk–bloem, vapour–steam, deel–flard

remaining three constant-language homogeneous sets and the corresponding heterogeneous ones, all pairs were in Dutch. In three of the homogeneous sets and the corresponding heterogeneous sets of the variable-language condition, two of the three prompt-response pairs were in English and the third pair was in Dutch. Thus, these sets have an English bias. In the remaining three variable-language homogeneous sets and the corresponding heterogeneous ones, two of the three prompt-response pairs were in Dutch and the third pair was in English. Thus, these sets have a Dutch bias. Regrouping the pairs of the constant-language sets created the variable-language sets. The third independent variable *language* refers to the actual language of the target words, Dutch or English.

In the homogeneous variable-language sets the first two segments were shared, but the remainder of the words was pronounced differently in the two languages. For example, the vowels in *blos*, *bloem*, *blade* are exclusive to the language of the words. Also, the phonological rules differ between

Dutch and English. For example, voiced consonants at the end of a word are devoiced in Dutch (i.e., the word-final /d/ becomes a /t/) but not in English. If a Dutch participant produces *blade* with its English pronunciation (i.e., English [bleɪd] rather than Dutch [bla.də]), then we know that the speaker used his or her uniquely English representations after the shared segments. If a preparation effect persists under a correct pronunciation of the words, then we have evidence for shared representations.

Each participant was tested once on each set. The order of the sets was fully counterbalanced across participants. Sixteen participants (groups A to D) were first tested on the homogeneous sets and then the heterogeneous ones. For the remaining 16 participants (groups E to H), the order of testing the homogeneous and heterogeneous conditions was reversed. The participants of groups A, B, E, and F were first tested on the constant-language sets, then on the variable-language ones. For participants of groups C, D, G, and H, the order of constant-language and variable-language sets was reversed. The participants of groups A, C, E, and G were first tested on the English-only and English-bias sets, then on the Dutch-only and Dutch-bias sets. For participants of groups B, D, F, and H, the order of testing was reversed. Each of the three prompt-response word pairs of a set was tested four times within each block of trials.

Results and discussion

Table 4 gives the mean production latencies and the error percentages for Experiment 2. The table lists for both the Dutch and English responses the means for the homogeneous and heterogeneous conditions when the

TABLE 4
Mean production latencies (M, in milliseconds), error percentages (E%), and preparation effects per blocking, language, and context for Experiment 2

		Context					
		Homogeneous		*Heterogeneous*		*Preparation*	
Blocking	*Language*	*M*	*E%*	*M*	*E%*	*M*	*E%*
Constant							
	English	724	1.6	749	2.2	–25	–0.6
	Dutch	711	2.6	739	2.9	–28	–0.3
	Mean	717	2.1	744	2.5	–27	–0.4
Variable							
	English	730	1.6	757	2.3	–27	–0.7
	Dutch	714	3.5	744	3.6	–30	–0.1
	Mean	722	2.5	751	3.0	–29	–0.5

languages were mixed (variable) and when they were not (constant). The table shows that a facilitatory effect from segmental overlap was obtained for both the constant-language and variable-language sets. For the variable-language sets with the Dutch bias, the preparation effect was 28 ms for the English words and 33 ms for the Dutch words in a set. For the variable-language sets with the English bias, the preparation effect was 24 ms for the Dutch words and 28 ms for the English words in a set. Thus, preparation is not confined to the two words sharing an onset in one language.

The production latencies were submitted to by-participant and by-item analyses of variance with the crossed variables context, blocking, and language. All variables were tested within participants. Context and blocking were tested within items and language was tested between items. There was a main effect of context, $F_1(1,31) = 13.00$, $MSE = 3770$, $p < .001$; $F_2(1,16) = 64.52$, $MSE = 214$, $p < .001$, but not of blocking, $F_1(1,31) < 1$, $F_2(1,16) = 5.76$, $MSE = 100$, $p < .03$, and not of language, $F_1(1,31) = 4.65$, $MSE = 2293$, $p < .04$; $F_2 < 1$. Most importantly, context and blocking did not interact, *F*s < 1. Also, there was no interaction between context and language, *F*s < 1, nor between blocking and language, *F*s < 1. Finally, there was no interaction between context, blocking, and language, *F*s < 1. Thus, the variable sets yielded a full-blown preparation effect, and the preparation effect was not restricted to words from one language. The analyses also confirm that bias had no effect. If preparation were dependent on bias, an interaction between context, blocking, and language should have been obtained.

These results show that a preparation effect is obtained when initial segments are shared even in the case that the language of the words in a set differs. Thus, speakers can plan initial segments common to both languages without knowing the language of the response word. This suggests that the representations of common segments are shared between the languages of bilinguals. If the segment representations were not shared between languages, the variable-language sets would be segmentally heterogeneous sets, and preparation should not have been obtained.

It may be that the preparation effect in variable sets is not due to representations shared between languages but that new segment representations were created in the course of the experiment. To perform the task, speakers might potentially create new within-language segment representations that are normally not used in that language. If participants are not using their natural segments but form new ones just for the task, then preparation effects should build up over time rather than be present from the outset. To examine the evolution of the preparation effect over trials, Experiment 3 was run. The experiment tested for a dependence of the preparation effect on response repetition. The new experiment included the variable-language sets of Experiment 2 only, and the number of times that each pair was tested within a block of trials was doubled, from four to eight.

EXPERIMENT 3

Method

Participants. The experiment was carried out with 12 Dutch participants with ages ranging from 18 to 29 years and who started to speak English at 11.6 years, on average.

Materials and design. The materials consisted of two practice sets and the twelve experimental sets of the variable-language condition of Experiment 2 (see Table 3). Thus, all variables of Experiment 2 were retained except blocking (constant, variable). Each set was tested in a separate block of trials. In six homogeneous sets, the response words shared the first two consonants (/st/, /bl/, and /fl/), and in the remaining six heterogeneous sets, they were unrelated in form. This is the first independent variable *context* (homogeneous, heterogeneous). The second independent variable is *language* (Dutch, English), referring to the actual language of the words. In three of the homogeneous sets and the corresponding heterogeneous ones, two of the three prompt-response pairs were in English and the third pair was in Dutch. In the remaining three homogeneous sets and the corresponding heterogeneous ones, two of the three prompt-response pairs were in Dutch and the third pair was in English. This third independent variable (not explicitly tested in Experiment 2) is referred to as *bias*.

Each participant was tested once on each set. The order of the sets was fully counterbalanced across participants. Each of the three prompt-response word pairs of a set was tested eight times within each block of trials. This fourth independent variable (not explicitly tested in Experiment 2) is called *repetition*.

Results and discussion

Table 5 gives the mean production latencies and the error percentages for context, language, and repetition in Experiment 3 (the means for bias are not shown). The table shows that a facilitatory effect from segmental overlap was obtained for both the Dutch and the English words, and that the size of the effects was similar for both languages. The production latencies were submitted to by-participant and by-item analyses of variance with the crossed variables context, bias, language, and repetition. All variables were tested within participants. Context and repetition were

TABLE 5
Mean production latencies (M, in milliseconds), error percentages (E%), and preparation effects per language and context for Experiment 3 (which included variable-language sets only)

		Context					
		Homogeneous		*Heterogeneous*		*Preparation*	
Language	*Repetition*	*M*	*E%*	*M*	*E%*	*M*	*E%*
English	1	725	2.8	759	9.3	–34	–6.5
	2	662	5.6	724	0.9	–62	4.7
	3	649	0.0	722	0.9	–73	–0.9
	4	658	3.7	710	1.9	–52	1.8
	5	663	5.6	711	0.9	–48	4.7
	6	630	0.9	716	2.8	–86	–1.9
	7	627	3.7	701	2.8	–74	0.9
	8	634	6.5	664	1.9	–30	4.6
	Mean	656	3.6	713	2.7	–57	0.9
Dutch	1	685	5.6	679	8.3	–12	–2.7
	2	674	4.6	706	3.7	–32	0.9
	3	653	2.8	716	2.8	–63	0.0
	4	631	2.8	679	0.9	–48	1.9
	5	634	4.6	664	3.7	–30	0.9
	6	613	4.6	676	2.8	–63	1.8
	7	610	3.7	644	4.6	–34	–0.9
	8	603	1.9	670	4.6	–67	–2.7
	Mean	638	3.8	682	3.9	–44	–0.1

tested within items, and bias and language between items. There was a main effect of context, $F_1(1, 11) = 20.87$, $MSE = 23941$, $p < .001$; $F_2(1, 14) = 43.57$, $MSE = 3785$, $p < .001$, but there were no reliable effects of bias, $F_1(1, 11) = 18.31$, $MSE = 6879$, $p < .001$; $F_2(1, 14) < 1$, and language, $F_1(1, 11) = 8.12$, $MSE = 6994$, $p < .02$; $F_2(1, 14) < 1$. There was no interaction between context and bias, $F_1(1, 11) = 2.84$, $MSE = 4954$, $p > .12$; $F_2(1, 14) = 2.04$, $MSE = 3785$, $p > .18$, nor between context and language, $F_1(1, 11) = 5.54$, $MSE = 3332$, $p < .04$; $F_2(1, 14) = 1.58$, $MSE = 3785$, $p > .23$, or between context, bias, and language, Fs < 1. Thus, preparation was again not restricted to words from one of the languages.

The latencies decreased with repetition, $F_1(7, 77) = 8.08$, $MSE = 6484$, $p < .001$; $F_2(7, 98) = 14.78$, $MSE = 1302$, $p < .001$, but there was no interaction between context, bias, and repetition, $F_1(7, 77) = 1.00$, $MSE = 4784$, $p > .44$; $F_2(7, 98) = 1.07$, $MSE = 1101$, $p > .39$, nor did repetition interact with any other variable or combination of variables, all ps $> .05$. In short, the preparation effect was present right from the start and did not develop over trials. This suggests that preparation is not due to new segment representations being created in the course of the experiment.

EXPERIMENT 4: ALL ENGLISH MATERIALS

The results of the previous two experiments showed that speakers can prepare segments such as /s/ and /t/ when the set includes, for example, English *steam* and *stone* and Dutch *stoel*. This suggests that the representations of segments common to the first and second language are shared between the languages in the memory of bilinguals. However, as indicated in the introduction section, it is important to determine the exact locus of the effect. Although preparation in the native language requires that the overlap in a set involves segments rather than features (Roelofs, 1999), it remains possible that preparation in the second language can be done in terms of features and that full segmental overlap is not necessary. Stemberger (1989) examined speech errors in natural speech and observed that young children are more likely to make phonological feature errors than adults are. Levelt et al. (1999) have suggested that features are chunked into segments only in the course of learning a language. So, it may be the case that second language segments (for unbalanced bilinguals) are represented in terms of their features, and that the preparation effect of shared segments observed in the previous experiments is due to preparation at the featural level.

Experiment 4 examined this possibility. The experiment tested whether preparation in a foreign language requires that the responses share initial segments or whether it suffices that they share features only. The experiment compared the preparation effect for sets whose words share initial segments such as *boat*, *bird*, and *boy* to the preparation effect for sets of words whose initial segments are the same except for one phonological feature, for example, *boat*, *bird*, and *pain*. In the latter set, /b/ and /p/ share all features except that /b/ is voiced and /p/ is voiceless. If a substantial preparation effect is obtained for the featural sets (e.g., *boat*, *bird*, *pain*), this implies that sharing full segments is not necessary for preparation in the second language, and the conclusion that representations of common segments are shared, drawn from Experiments 2 and 3, would be undermined. If features suffice, the effect in the variable-language condition of Experiments 2 and 3 may have arisen from preparing the first language words at the segmental level and the second language words at the featural level.

WEAVER++ predicts that feature overlap alone should yield no preparation effect. In the model, the preparation effect results from the seriality of phonological encoding (Roelofs, 1999). The features of a string of segments may be accessed in parallel during phonetic encoding. Preparing features therefore will not yield facilitation. Thus, WEAVER++ predicts that the responses should be faster in the homogeneous condition with the segment overlap than in the three other conditions (segment heterogeneous, feature homogeneous, and feature heterogeneous), which should not differ.

Method

Participants. The experiment was carried out with eight Dutch participants with ages ranging from 18 to 33 years and who started to speak English at 11.8 years, on average.

Materials and design. The materials consisted of two practice sets and twenty-four experimental sets of three prompt-response word pairs each (see Table 6). The prompts and responses were in English. Each set was tested in a separate block of trials. There were 12 homogeneous sets and 12 heterogeneous ones. In six homogeneous sets, the response words shared the first consonant, and in the six other homogeneous sets the first segment shared all features except one. This independent variable, which had two levels (segments, features), will be called *level of overlap*.

TABLE 6
Prompt-response pairs of Experiment 4

Level of overlap	*Context*	*Set*
Segments	Homogeneous	Set 1: river–boat, wing–bird, girl–boy
		Set 2: night–day, cat–dog, window–door
		Set 3: opinion–view, rose–vase, speech–voice
		Set 4: war–peace, colour–paint, smoke–pipe
		Set 5: village–town, sugar–tea, clock–time
		Set 6: quarrel–fight, sea–fish, head–face
Segments	Heterogeneous	Set 7: cat–dog, speech–voice, river–boat
		Set 8: opinion–view, wing–bird, night–day
		Set 9: girl–boy, window–door, rose–vase
		Set 10: clock–time, sea–fish, war–peace
		Set 11: head–face, smoke–pipe, village–town
		Set 12: colour–paint, sugar–tea, quarrel–fight
Features	Homogeneous	Set 13: wing–bird, colour–paint, girl–boy
		Set 14: cat–dog, night–day, sugar–tea
		Set 15: head–face, speech–voice, opinion–view
		Set 16: war–peace, river–boat, smoke–pipe
		Set 17: clock–time, village–town, window–door
		Set 18: sea–fish, quarrel–fight, rose–vase
Features	Heterogeneous	Set 19: girl–boy, cat–dog, head–face
		Set 20: speech–voice, colour–paint, night–day
		Set 21: sugar–tea, opinion–view, wing–bird
		Set 22: smoke–pipe, village–town, rose–vase
		Set 23: quarrel–fight, river–boat, clock–time
		Set 24: window–door, sea–fish, war–peace

The critical segments in the experiments were the minimal pairs /d/ (voiced) and /t/ (voiceless), /b/ (voiced) and /p/ (voiceless), and /v/ (voiced) and /f/ (voiceless). All responses occurred in all conditions. In three homogeneous and the corresponding heterogeneous sets of the segment condition, the first segment was the voiced member of a minimal pair. In the remaining three homogeneous and heterogeneous sets of the segment condition, the first segment was the voiceless member of a minimal pair. In three homogeneous and the corresponding heterogeneous sets of the feature condition, two of the responses had the voiced member of a minimal pair as first segment and the third response had the voiceless member as first segment. In the remaining three homogeneous and heterogeneous sets, two of the responses started with the voiceless member and the third response started with the voiced member.

Each participant was tested once on each set. The order of the sets was fully counterbalanced across participants. Four participants (groups A and B) were first tested on the homogeneous sets, and then on the heterogeneous ones. For the remaining four participants (groups C and D), the order of testing the homogeneous and heterogeneous conditions was reversed. The participants of groups A and C were first tested on the sets of the segment condition, then on the sets of the feature condition. For participants of groups B and D, the order of testing these conditions was reversed. Each of the three prompt-response word pairs of a set was tested four times within each block of trials.

Results and discussion

Table 7 gives the mean production latencies and the error percentages for Experiment 4. The table shows that the production latencies were faster for the condition in which the segments were shared than for the other three conditions, which did not differ much. Thus, a facilitatory effect from homogeneity was obtained for the sets with the segment overlap, but there was no effect at all for the sets with the feature overlap. The latencies were

TABLE 7
Mean production latencies (M, in milliseconds), error percentages (E%), and preparation effects per level of overlap and context for Experiment 4

	Context					
	Homogeneous		*Heterogeneous*		*Preparation*	
Level of overlap	*M*	*E%*	*M*	*E%*	*M*	*E%*
Segments	674	4.5	712	4.2	−38	0.3
Features	716	6.1	716	5.0	0	1.1

submitted to by-participant and by-item analyses of variance with the crossed variables context and level of overlap. Both variables were tested within participants and within items. The interaction between context and level of overlap was significant, $F_1(1, 7) = 17.10$, $MSE = 169$, $p < .001$; $F_2(1, 17) = 10.98$, $MSE = 591$, $p < .001$. The simple effect of context for segment overlap was significant, $F_1(1, 7) = 11.32$, $MSE = 507$, $p < .01$; $F_2(1, 17) = 12.57$, $MSE = 1028$, $p < .001$, but there was no effect for feature overlap, *F*s < 1. Main effects of context and level of overlap were obtained by items only, respectively, $F_1(1, 7) = 2.92$, $MSE = 979$, $p > .13$; $F_2(1, 17) = 4.39$, $MSE = 1466$, $p < .05$, and $F_1(1, 7) = 1.35$, $MSE = 2992$, $p > .28$; $F_2(1, 17) = 13.38$, $MSE = 681$, $p < .001$.

These results show that a preparation effect in the second language is only obtained when words share an initial segment, but not when all but one of the features of the first segment are shared. Thus, the finding that preparing the production of first language words requires shared segments rather than features was replicated for the production of second language words. Apparently, preparation in both the first and second language requires that the words in a set have segmental overlap. Thus, the conclusion from Experiments 2 and 3 that the representations of common segments (rather than features) are shared between languages remains supported. Furthermore, the results exclude that the preparation effect in Experiments 2 and 3 was due to articulatory preparation, that is, moving the articulators in the correct starting position before the beginning of a trial. The critical segments in the homogeneous condition of the current experiment always shared the place and manner of articulation and differed only in voicing (voiced versus voiceless). Thus, if articulatory preparation were the source of the preparation effects in the preparation paradigm, facilitation should have been obtained in both the segment and feature condition. However, since a preparation effect is observed only when segments are shared but not when features are shared, the locus of the preparation effect must be the level of phonological encoding rather than a lower level, such as phonetic encoding or articulatory preparation. Of course, only the feature of voicing was tested in Experiment 4, but Roelofs (1999) showed that the same holds for other features in first language production.

GENERAL DISCUSSION

The present study examined two aspects of the process of phonological encoding in a second language. First, several studies have shown that the process of phonological encoding in the first language proceeds from the beginning of a word to its end (Meyer, 1990, 1991; Meyer & Schriefers, 1991; Roelofs, 1996a, 1998a). It was an open issue whether this also holds for second language words. Dell et al. (1997) have provided evidence that when serial order behaviour becomes more practiced, the serial ordering mechanism becomes more efficient in controlling linear order. Hence, it may be that linear order is highly controlled in producing first language words (hence no effect of end overlap) but not (yet) in producing second language words. Experiment 1 showed, however, that the seriality effect that is normally observed for producing first language words is replicated for producing words in a second language. Second, the present study examined whether the representations of segments that are common to the first and second language (Ladefoged, 2001; Ladefoged & Maddieson, 1996) are shared between languages in memory. The findings from Experiments 2 and 3 suggest that speakers may construct part of the form plan of words without knowing in advance the language. This suggests that at least some of the representations recovered from memory are shared between languages. Experiments 2 and 3 left open whether there are shared representations of segments or features. The results of Experiment 4 suggest that segments are shared rather than their features only. To summarise, the results from Experiments 1 to 4 show that even bilinguals who are relatively fluent but unbalanced across the two languages show strikingly similar preparation patterns for their two languages. This suggests that bilingual speakers can be functionally monolingual as far as rightward incrementality and the representation of common segments is concerned. Whenever possible, word forms in both languages seem to be ‘‘woven on the same loom’’.

The Dutch population tested in the current study was fluent in English, but an average starting age of more than 11 years means that they would classify as late bilinguals. It may be that for truly early (i.e., ‘‘balanced’’) bilinguals, separate phonological systems might be more plausible than for the tested population (e.g., MacSwan, 2000). For early bilinguals, the phonological space is probably more distinctly defined than for late bilinguals and hence the languages could be more clearly separated. Furthermore, English and Dutch are two languages that largely overlap in their phonological space. The current results do not neccessarily hold true for languages that are more clearly separated in their respective sound contrasts. Whether the degree of language skills and type of languages of a bilingual play a role may be examined in future studies.

The current research casts doubt on the suggestion of Kim et al. (1997) that there are separate representations of the sound structures of the languages of late bilinguals. In their fMRI study, bilinguals silently described what they had done during the previous day. In late bilinguals, the activation during second language use was spatially separated from the activation during first language use in Broca’s area. However, comparing discourses in two languages does not allow a distinction between phonological and phonetic levels, or between these and any other levels. The current results suggest that phonological segments and their serial ordering are shared between languages in late bilinguals, and therefore cannot have caused the spatially separate brain activations.

In examining bilingual speech errors or conducting bilingual picture-word interference experiments, where spoken primes in one language are presented during picture naming in another language (Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, de Bot, & Schreuder, 1998), it is difficult to differentiate between shared representations or closely linked ones. Therefore, I used an on-line form-preparation paradigm in all experiments. Obviously, this experimental task differs in a number of ways from ordinary speech production. Speakers rarely say the same three words over and over again, and they cannot normally predict how the next word to be uttered will begin. Yet, the preparation effects in the form-preparation experiments show very systematic patterns. Speakers can exploit certain types of information, whereas others are completely useless. A natural account of these patterns is to relate them to the way speech is normally planned. Thus, I assume that the reason why preparation effects were only obtained for initial overlap is that in normal bilingual speech production the planning proceeds from the beginning of an utterance to its end. Similarly, it is plausible to assume that participants can prepare the initial segments common to both languages without knowing the language of the word because these segments are shared between the languages in normal language production. And sharing features alone yields no preparation, because features are accessed in parallel from segments in normal bilingual production.

Although the present findings may be explained in terms of the mechanisms of normal language production, it remains possible that participants have adopted a special strategy in the experiments. For example, if segments are not shared between languages, participants may adopt the strategy of using first-language segments (e.g., Dutch /s/ and /t/) in planning the second-language words in the homogeneous sets of the variable-language condition (or second-language segments in producing first-language words). However, since this strategy would imply a type of planning that is not present in normal production, namely one that includes within-word language switching, one would expect that a cost is associated to it. The words in the variable-language condition, requiring this unusual type of planning, would be expected to have longer production latencies than the same words in the constant-language condition. This, however, appears not to be the case. In Experiment 2, the mean production latencies of the words in the homogeneous constant-language and variable-language conditions were 717 and 722 msec, respectively, with Fs < 1. So, preparation in the homogeneous variable-language condition implies no extra cost, which supports the idea that the words are planned the same in all conditions.

A related strategy might be that speakers are simply producing the odd word in a set as if it were in the same language as the other words. For example, the English word *blade* in the set *blos*, *bloem*, *blade* may be pronounced as a Dutch non-word. However, in the homogeneous variable-language sets the first two segments were shared but the remainder of the words was pronounced differently in the two languages. Since the participants produced the English words with their English pronunciation and the Dutch words with their Dutch pronunciation, we know that the participants used their uniquely English and Dutch representations after the shared segments. That is, a word like *blade* was produced with its English pronunciation [bleɪd] rather than as a Dutch non-word, namely disyllabic [bla.də]. This rules out that the participants treated the English words as Dutch non-words or the Dutch words as English non-words. Because the preparation effect persisted under a correct pronunciation of the words, we have evidence for shared representations.

A final strategy of participants might have been to create *new* within-language segments of sounds that are normally not in that language. For example, for the set *blos*, *bloem*, *blade*, they might create the segment string /bleɪd/ for Dutch that sounds like the English word *blade* (thereby modifying the Dutch language). However, if the preparation effect were due to such new within-language segments that are created in the course of the experiment, the preparation effect should have developed over trials (learning new sounds in a language takes time). But Experiment 3 showed that the preparation effect was present right from the start and did not develop over trials. Furthermore, in generating the new Dutch segment string /bleɪd/, the regular Dutch pronunciation /bla.də/ of the non-word *blade* should be prevented. Thus, one expects that a cost is associated with this unusual type of planning that involves blocking a regular pronunciation. However, as indicated above, the constant-language and the variable-language words were produced with latencies that were statistically the same. So, there is no support for the assumption that the preparation effect is due to new segment representations created in the course of the experiment.

An alternative explanation for the current results would be that the cross-language preparation effect is based on orthography. This would explain the existence of facilitation in the mixed language conditions and the absence of facilitation when features, but not phonological segments are shared. However, earlier work has suggested that preparation effects occur even when the orthography is not shared. Meyer (1990) observed preparation effects for shared syllables (i.e., [si:]) of 42 ms (her Experiment 1) and 49 ms (her Experiment 3) with varying orthography (i.e., *ci* versus *si*). Similarly, Chen, Chen, and Dell (2002) demonstrated in Mandarin Chinese that preparation of syllables is not dependent on shared orthography by using two-syllable two-character response words. When the first syllable but not the character is shared, the preparation effect is as large as when the syllable and its character are shared. These results suggest that orthography does not play a role in the preparation task.

The findings about phonological encoding in the present paper support Grosjean's (1982) case against considering bilinguals as two monolinguals in one body. This does not exclude, of course, that at some levels of their languages they may be functionally two monolinguals. For example, an English–Finnish bilingual has probably different procedures for morphological encoding in the two languages, which is agglutinative for Finnish and non-agglutinative for English. However, the current findings suggest that a basic rightward incremental mechanism for planning the phonological shape of utterances in production is shared between the first and second language of a bilingual, and that the mechanism may use shared representations of segments common to the languages.

Manuscript received November 2001
Revised manuscript received July 2002

REFERENCES

Albert, M., & Obler, L.K. (1978). *The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism*. London: Academic Press.

Altenberg, E., & Cairns, H. (1983). The effects of phonotactic constraints in lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior, 22*, 174–188.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database*. (CD-ROM). Linguistic Data Consortium, University of Pennsylvania; Philadelphia, PA.

Baddeley, A.D. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.

Booij, G.E. (1995). *The phonology of Dutch*. Oxford: Oxford University Press.

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society, 13*, 212–214.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751–781.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41*, 365–397.

De Groot, A.M.B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–51). Amsterdam: John Benjamins Publishing Company.

De Groot, A.M.B., & Kroll, J. (Eds.) (1997). *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahaw, NJ: Lawrence Erlbaum Associates Inc.

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G.S., Burger, L.K., & Sved, W.R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123–147.

Donders, F. C. (1868). Over de snelheid van psychische processen. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868–1869, Tweede reeks*, *II*, 92–120. Reprinted as Donders, F.C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Ervin, S., & Osgood, C. (1954). Second language learning and bilingualism. *Journal of Abnormal and Social Psychology*, *49*, 134–146.

Flege, J.E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395–441.

Flege, J.E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*, 185–202.

Flege, J.E., & Port, R. (1981). Cross-language phonetic interference. *Language and Speech, 24*, 125–146.

Francis, W. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, *125*, 193–222.

Grainger, J. (1993). Visual word recognition in bilinguals. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 11–25). Amsterdam: John Benjamins Publishing Company.

Grainger, J. & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R.J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 207–220). Amsterdam: North Holland.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.

Hermans, D., Bongaerts, Th., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*, 213–229.

Indefrey, P., & Levelt, W.J.M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: MIT Press.

Jared, D., & Kroll, J.F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*, 2–31.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Oxford: Blackwell Publishers.

Kim, K.H.S., Relkin, N.R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*, 171–174.

Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.

Kroll, J.F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 53–81). Amsterdam: John Benjamins Publishing Company.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sound of languages.* Oxford: Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishers.

Leopold, W. (1939). *Speech development of a bilingual child: A linguist's record. Volume I: Vocabulary growth in the first two years.* New York: AMS Press. Reprinted in 1970.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1–22.

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

Mackey, W. (1967). *Bilingualism as a world problem*. Montreal: Harvest House.

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, *3*, 37–54.

Meyer, A.S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545.

Meyer, A.S. (1991). The time course of phonological encoding in language production: The phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89.

Meyer, A.S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1146–1160.

Paradis, M. (1995). *Aspects of bilingual aphasia*. Oxford: Pergamon Press.

Perani, D., Paulesu, E., Sebastian Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841–1852.

Poulisse, N. (1990). *The use of compensatory strategies by Dutch learners of English*. Dordrecht: Mouton De Gruyter.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Amsterdam: John Benjamins.

Pullum, G., & Ladusaw, W.A. (1996). *Phonetic symbol guide*. Chicago: University of Chicago Press.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.

Roelofs, A. (1994). On-line versus off-line priming of word-form encoding in spoken word production. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 772–777). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Roelofs, A. (1996a). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854–876.

Roelofs, A. (1996b). Morpheme frequency in speech production: Testing WEAVER. In G.E. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1996* (pp. 135–154). Dordrecht: Kluwer Academic Publishers.

Roelofs, A. (1997a). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (1997b). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12*, 659–696.

Roelofs, A. (1998a). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904–921.

Roelofs, A. (1998b). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition*, *1*, 94–95.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14,* 173–200.

Roelofs, A. (2000). Word meanings and concepts: What do the findings from aphasia and language specificity really say? *Bilingualism: Language and Cognition*, *3*, 25–27.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

Roelofs, A., & Meyer, A.S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939.

Schreuder, R., & Weltens, B. (Eds.) (1993), *The bilingual lexicon*. Amsterdam: John Benjamins Publishing Company.

Snodgrass, J. G. (1993). Translating versus picture naming: Similarities and differences. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 83–114). Amsterdam: John Benjamins Publishing Company.

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, *28*, 164–188.

Watson, I. (1991). Phonological processing in two languages. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 25–48). Cambridge: Cambridge University Press.

Weinreich, U. (1953). *Languages in contact: Findings and problems*. New York: Linguistic Circle of New York. Reprinted in 1974 by Mouton, The Hague.