

# The gender marking effect in spoken word recognition: The case of bilinguals

DELPHINE GUILLELMO and FRANÇOIS GROSJEAN  
*University of Neuchâtel, Neuchâtel, Switzerland*

It has been known for some time that the recognition of a noun is affected by the gender marking, such as masculine or feminine, that is carried by a preceding word. In this study, we used auditory naming to examine how early and late English–French bilinguals react to gender marking when processing French. The early bilinguals showed clear facilitation and inhibition effects, but the late bilinguals were totally insensitive to gender marking, whether congruent or incongruent. The results are discussed in terms of current accounts of gender processing as well as age of acquisition and regular use of the gender-marking language.

There is increasing evidence that in languages that have gender agreement, a congruent gender marking usually speeds up the processing of the following noun relative to an incongruent marking (or no marking). This effect is now well established in monolinguals, but little is known about how bilinguals react to gender agreement. In this paper, we ask whether bilinguals show the same effect and whether it depends on when they acquired and started using the gender-marking language on a regular basis.

In what is fast becoming a classic, Corbett (1991) stated that gender is the most puzzling of the grammatical categories that interests nonlinguists as well as linguists, and that it becomes more fascinating the more it is investigated. Gender can be defined as follows: “A subclass within a grammatical class (as noun, pronoun, adjective, or verb) of a language that is partly arbitrary but also partly based on distinguishable characteristics (as shape, social rank, manner of existence, or sex) and that determines agreement with and selection of other words or grammatical forms” (Webster’s Ninth New Collegiate Dictionary, 1991). Depending on the language, words (nouns usually) carry any number of genders; from two in such languages as Italian, Spanish, and French, all the way to six for Swahili. Of particular interest here is that other word classes in a language that has gender, such as adjectives, verbs, articles, pro-

nouns, and so on, do not have a gender per se but can reflect, in their inflectional morphology, the gender of the words that do. Thus, depending on the language, a gender agreement marking can appear before or after a noun on a determiner, adjective, pronoun, and so on. In the case of French, for example, *voiture* is feminine, and in the phrase *leur petite voiture* (*their small car*), the adjective (*petite*) agrees with the noun and carries a feminine ending. In the phrase *le garçon*, the definite article is masculine since the noun is of that gender.

It is now well established that a congruent gender marking will speed up the processing of the following noun. This has been shown in reading, for example, by Gurjanov, Lukatela, Lukatela, Savic, and Turvey (1985) in Serbo-Croatian and by Colé and Segui (1994) in French. In speech, Grosjean, Dommergues, Cornu, Guillelmo, and Besson (1994), working with French, showed with a gating task that participants needed less of a noun to identify it when it was preceded by a gender-congruent article (they were also more confident about the word they proposed), and with a lexical decision task, they showed that the participants were faster at deciding that the noun was a word. Recently, Jakubowicz and Faussart (1998), also working on spoken French and using a lexical decision task, replicated a strong gender effect.

Researchers have also used a neutral or baseline condition in order to determine whether the effect is due to congruency (facilitation), incongruency (inhibition), or both. Schmidt (1986), for example, found a significant incongruency effect in German, but not a congruency effect (although there was a trend in the right direction). In the spoken modality, Bates, Devescovi, Hernandez, and Piz-zamiglio (1996), working with Italian, found both an incongruency (inhibition) effect and a congruency (facilitation) effect with an auditory naming task (also called cued shadowing; Bates & Liu, 1997) but only an incongruency effect with a gender-monitoring task. They concluded that the inhibitory component of gender priming is more robust than the facilitatory component.

---

Preparation of this paper was made possible in part by two grants from the Swiss National Science Foundation (1213-045375.95 and 1214-058848.99). We thank the monolingual and bilingual participants who took part in the five experiments, as well as Harlan Lane, Joanne Miller, and Neal Pearlmutter for their valuable comments during an oral presentation of the data. Additional thanks go to Lysiane Grosjean, Nicolas Léwy, Isabelle Racine, and Carole Yersin-Besson for their careful reading of the manuscript. Special thanks go to Bob Lorch and to two anonymous reviewers for their very useful suggestions during the revision process. Correspondence should be addressed to F. Grosjean, Laboratoire de traitement du langage et de la parole, Université de Neuchâtel, Avenue du Premier-Mars 26, 2000 Neuchâtel, Switzerland (e-mail: francois.grosjean@unine.ch).

In regard to the locus of the gender-marking effect, Grosjean et al. (1994) mentioned two possibilities. On the one hand, the process could be internal to the lexicon in that when a word carrying a gender marking is identified, it activates all the nouns in the lexicon that share the same gender. A variant of this first possibility is that it is the gender feature of the word that activates all the nouns with that same feature. Whatever the variant, this first explanation could account for faster processing in the congruent condition (i.e., when the preceding word and the noun share the same gender) and slower processing in the incongruent condition, whatever the task used. An alternative explanation involves both the lexical and syntactic processing modules.<sup>1</sup> The lexical module undertakes the recognition of the words in question (e.g., an article and a noun), and the syntactic module checks that gender agreement is respected (for a similar proposal, see, among others, Gollan & Frost, 2001; Van Berkum, 1996). Along this line, Jakubowicz and Faussart (1998) stated that the gender-marking word “sets” the feature value for the entire phrase, and the congruency and incongruency effects are the result of an automatic post-access check of the grammatical agreement between the words that have a gender marking. This checking mechanism will speed up the participant’s task when there is congruency and slow it down when there is incongruency. Recently, in a paper dealing with another agreement phenomenon (number), Pearlmutter, Garnsey, and Bock (1999) further elaborated on this grammatical explanation by stating that there might be two processing possibilities: a compute-on-the-fly system where the agreement features are processed by the comprehension system as they are encountered and a backtrack mechanism where agreement is checked after the initial parsing, and only when possible (i.e., when the word in question is overtly marked for the agreement feature). Grosjean et al. stated that it might well turn out that the gender-marking effect is both a lexical effect *and* a syntactic effect, and they provided evidence for this. Bates, Devescovi, Pizzamiglio, d’Amico, and Hernandez (1995) also stated that gender congruency effects may well involve a combination of lexical (they call it “prelexical”) and syntactic (“postlexical”) processing.

However one may explain the gender-marking effect in the end, it should be noted that it has been studied primarily with monolinguals. It is therefore normal to ask whether bilinguals will show the same effect and whether it depends on when they acquired and started using the gender-marking language in question on a regular basis. At first sight, it seems natural to predict that bilinguals should show the effect. After all, gender marking can be useful for a number of things irrespective of whether the listener is monolingual or bilingual. First, it may preactivate a class of nouns and/or be used in a post-access agreement check (as discussed above). Second, it can help re-group words into phrases (Van Berkum, 1996). And third, at the level of discourse processing, gender marking can help keep track of referents and so help disambiguate

anaphoric or deictic referential constructions (Cacciari, Carreiras, & Barbolini Cionini, 1997; Corbett, 1991). This said, it might also be the case that the presence of a gender-marking effect in bilinguals may depend on how old they were when they started acquiring and using the language(s). Second-language acquisition research seems to show that early bilinguals (i.e., those who acquired and used their gender-marking language regularly before adolescence) make no, or very few, gender-production errors, whereas late bilinguals (i.e., those who acquired their other language during adolescence or as adults) make a substantial number of gender errors (see, e.g., Carroll, 1989; Rogers, 1987). This is reflected anecdotally in an interview given by Sir Winston Churchill on French radio in 1946. He was answering questions in fluent French accompanied by a heavy English accent. At one point he stated with humor: “Despite working so hard and coming so far with the French to help them win their freedom, I have never mastered the gender of French nouns!”

The question that one can ask is whether perception will parallel production—that is, whether early bilinguals who make no gender errors are sensitive to gender marking when processing language, whereas late bilinguals are less sensitive to it. Alternatively, the perception and production of gender marking might behave differently in the sense that, irrespective of what happens in production, both types of bilinguals are sensitive to it to the same extent in perception. This paper will report two experiments. In the first experiment, we will show how early English–French bilinguals react to gender marking when processing French, and in the second experiment we will examine how late bilinguals do so. In both cases, the results of the bilinguals will be compared with those of French monolinguals.

## **EXPERIMENT 1** **Early Bilinguals**

The aim of this experiment was twofold. First, we wished to replicate with French monolinguals the results of the Bates et al. (1996) study, which showed both a congruency (facilitation) effect and an incongruency (inhibition) effect with auditory naming. Although the language used here is different (they used Italian), the gender system is highly similar in the two languages, and there is no a priori reason to believe that the two effects cannot be replicated. Since a pilot study showed that the strength of the congruency effect depends, in part at least, on the grouping of the experimental conditions, we opted for two groupings: congruent and neutral, and incongruent and neutral. Participants heard short noun phrases made up of a determiner, an adjective, and a noun, and they were asked to repeat the noun.<sup>2</sup> Depending on the part of the study, the gender marking on the determiner was either congruent or neutral, or incongruent or neutral, with respect to the noun. Our second aim was to assess whether early bilinguals were sensitive to gender marking. They

**Table 1**  
**Biographical and Language Proficiency Means**  
**and Standard Deviations for the Two Groups of**  
**English–French Bilinguals ( $n = 32$  in Each Group)**

Biographical and Language Proficiency Categories	Early Bilinguals		Late Bilinguals	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	24;4	6;6	48;5	10;6
Age started speaking English	1;11	1;4	1;0	0;0
Age started speaking French	4;0	3;0	15;11	7;11
Age of onset of bilingualism	5;4	3;5	24;8	6;0
English oral comprehension*	6.4	0.6	6.8	0.4
French oral comprehension*	6.7	0.5	6.0	0.8
English oral production*	5.8	0.8	6.7	0.5
French oral production*	6.3	0.7	5.4	1.0

\*Based on a self-rating scale (1 = *very poor*; 7 = *excellent*).

had acquired English and French in their early childhood, and so it was expected that their behavior would be similar to that of their monolingual counterparts.

## Method

**Participants.** Two groups of participants were used in this study. The first was made up of 32 native monolingual French-speaking students of the University of Neuchâtel (Switzerland), with no reported speech or hearing defects, who served individually in the experiment. They were assigned at random to one of two experimental subgroups of 16. The second group was made up of an equal number of early English–French bilingual students, with no reported speech or hearing defects, who also served individually in the experiment. To be included in this “early bilingual” group, a participant had to report having started using both languages on a regular basis in childhood (the maximum limit for this “onset of bilingualism” was 13 years old). Most participants were well under this limit; on average, they started using their two languages in everyday life as early as 5;4 years. A questionnaire was filled out by the bilingual participants, and relevant biographical and language proficiency data are summarized in Table 1 (left columns). We can observe that the bilinguals were relatively young (24;4 years on average) and that they started speaking English about 2 years before they did French (1;11 and 4;0 years, respectively). There was a slight gap in time between starting to speak the second language (4;0 years) and using the two languages on a regular basis (i.e., age of onset of bilingualism: 5;4 years). This probably reflects the difference between episodic use and regular use of the second language. In regard to reported language proficiency obtained with a self-rating scale (1 = *very poor*; 7 = *excellent*), the participants reported very high levels of spoken English and spoken French comprehension (6.4 and 6.7, respectively). Their reported levels of spoken language production were practically as high (5.8 and 6.3, respectively). The bilinguals showed a slight dominance in French, but it was quite small, especially in spoken comprehension, the skill that interests us here (difference of 0.3 between the two languages). The bilingual participants were also assigned at random to one of two experimental subgroups of 16.

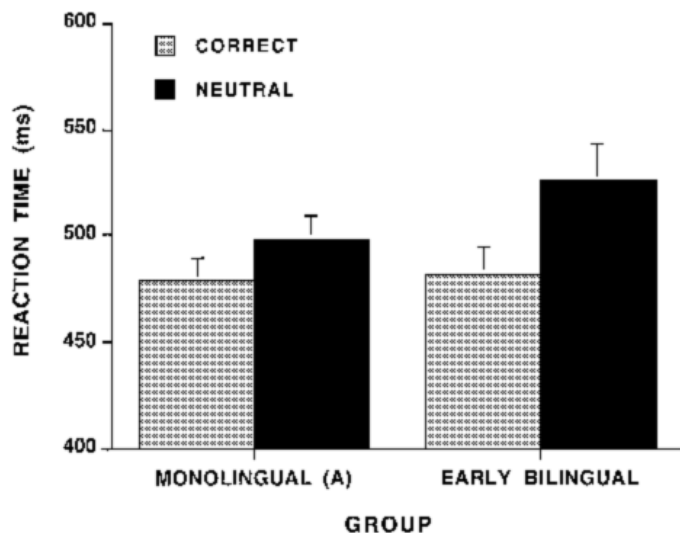
**Materials.** Thirty-six French nouns—18 masculine and 18 feminine—were chosen for the study (see the Appendix for a complete list). All of them started with a stop consonant. Half the words in each set were one syllable long and the other half were two syllables long. Care was taken to make sure that the two sets of nouns had the same mean frequency of occurrence: 2,607 for the masculine words and 2,502 for the feminine words, based on the BRULEX database (Content, Mousty, & Radeau, 1990). Their uniqueness points (UPs) were also similar: 11 masculine words and 12 feminine words had a UP after the end of the word whereas 7 masculine words and 6 feminine words had it before the end (based

on *Le Robert Oral–Écrit*, 1989). All UPs were the same whether the words were preceded by a correct, an incorrect, or a neutral gender marking. (It should be noted that Grosjean et al., 1994, had a similar set of one- and two-syllable words, half masculine and half feminine; they showed that the gender-marking effect is very robust and is not affected by differences in the length and in the gender of the words used.)

The stimuli were prepared in three stages. In the first stage, three determiners were chosen: *le* (masculine *the*), *la* (feminine *the*), and *leur* (neutral *their*). The three were read 20 times in a short context: *le coq* (*the rooster*), *la coque* (*the hull*), and *leur coq/coque* (*their rooster/hull*). (The pronunciation of *coq* and *coque* is identical in French despite the difference in orthography.) The recording took place in a sound-proof studio, and a native speaker of French read the phrases at a normal rate. The recording was then digitized on a Macintosh II at a sampling rate of 22 kHz with the Sound Designer II package. The five best exemplars of each determiner, *le*, *la*, and *leur*, on the basis of an evaluation by five judges, were then spliced out. These were measured and a final exemplar of each was chosen so that its duration was similar to that of the other two (173 msec for *le*, 174 msec for *la*, and 173 msec for *leur*). In the second stage, each of the 36 nouns were read by the same speaker in three contexts: *le joli* \_\_\_\_\_ (*the nice* \_\_\_\_\_), *la jolie* \_\_\_\_\_, and *leur joli* \_\_\_\_\_. (It should be noted that the acoustic characteristics of *joli(e)* are not changed when preceded by a masculine, feminine, or neutral determiner.) Following this, one third of the readings in each context was retained so that each noun appeared once. For example, if the reading *la jolie glace* (*the nice mirror*) was retained, the other two readings (*\*le joli glace* and *leur jolie glace*) were discarded. Thus, a third of the adjective + noun pairs retained came from the congruent reading context, a third from the incongruent context, and the final third from the neutral context. The determiners from each reading were then spliced out so as to leave 36 *joli(e)* + noun segments, one for each of the 36 nouns. Finally, in the last stage, each experimental determiner obtained in Stage 1 (*le*, *la*, *leur*) was added to each *joli(e)* + noun segment to give three experimental exemplars: *le* + *joli* + noun (where the gender marking is correct if the noun is masculine and incorrect if it is feminine), *la* + *jolie* + noun (where the gender marking is correct if the noun is feminine and incorrect if it is masculine), and *leur* + *joli(e)* + noun (where the gender marking is neutral since *leur* carries no gender information). Thus, for example, the three experimental exemplars for *bateau* (boat) were *le joli bateau* (here the gender marking is correct), *\*la jolie(e) bateau* (the gender marking is incorrect), and *leur joli bateau* (the gender marking is neutral). A short (1000-Hz) tone was placed at the onset of the noun on the right channel for timing purposes, and the interstimulus interval (ISI) was set at 3.5 sec.

Two groups of experimental stimuli were prepared; one group contained the correct and neutral stimuli (for the correct/neutral part of the study), and the other group contained the incorrect and neutral stimuli (for the incorrect/neutral part). Two sets of stimuli were then prepared for each part, each set containing 18 correct stimuli (or incorrect stimuli, depending on the part) and 18 neutral stimuli. A particular noun appeared only once in each set, preceded by the correct (or incorrect) gender marking or preceded by the neutral gender marking. The order of presentation of the stimuli in each set (correct or incorrect, and neutral stimuli) was the same and was quasi-random.

**Procedure.** The experiment was run with PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), and participants were tested individually in a quiet environment. Each participant, monolingual or bilingual, was tested on only one part—either the correct/neutral part or the incorrect/neutral part—and on only one set of stimuli within each part. Participants were asked to listen to the phrases presented to them over headphones and to repeat the word after *joli(e)* as quickly as possible. Reaction times were recorded by means of a Hewlett-Packard universal counter (HP 5315) that was started by



**Figure 1.** Mean naming times (in milliseconds) as a function of group (monolingual [A] and early bilingual) and gender marking (correct and neutral). Error bars represent  $+1$  SEM.

the tone placed on the right channel (at the onset of each noun) and stopped by the participant's vocal response.<sup>3</sup> A short practice session took place before the experimental session and a short break occurred half way through the experiment. French was used throughout the testing session, which was conducted by a native speaker of French (D.G.).

## Results and Discussion

The following data analysis procedure was applied to the reaction times obtained. First, times above a particular cutoff point (1,000 msec) were removed. Then, for each participant, an overall mean and standard deviation was calculated, and the values above or below  $2$  SD were replaced with values that corresponded to the mean plus  $2$  SD, or to the mean minus  $2$  SD, respectively. The values that were removed because they were greater than 1,000 msec were then replaced with the new mean, as were the missing values. For the monolinguals, in the correct/neutral part there were no times greater than 1,000 msec, 4.17% of the values were above or below  $2$  SD, and 3.30% of the values were missing. In the incorrect/neutral part, 0.17% of the times were greater than 1,000 msec (and hence replaced by the new mean), 2.95% of the values were above or below  $2$  SD, and 2.78% of the values were missing. For the early bilinguals, in the correct/neutral part there were no times greater than 1,000 msec, 3.99% of the values were above or below  $2$  SD, and 2.08% of the values were missing. In the incorrect/neutral part, there were no times greater than 1,000 msec, 5.03% of the values were above or below  $2$  SD, and 1.74% of the values were missing. Two-way analyses of variance (ANOVAs), over participants and over items, were conducted for each part: correct and neutral, and incorrect and neutral. For all of these analyses, alpha was set at .05.

Figure 1 presents the results of the correct/neutral part of the study. Mean naming times (in milliseconds) are plotted as a function of group—the monolingual group, (la-

beled “Monolingual (A)”) and the early bilingual group—and gender marking—correct and neutral. Henceforth, the neutral condition will always appear as a black bar. As can be seen, a congruency effect was present for both the monolinguals and the early bilinguals, but it was larger for the latter. When a noun was preceded by a determiner that marked the gender, monolingual participants needed 479 msec on average to name a noun preceded by a correct gender marking (e.g., *le joli bateau*) and 498 msec to name it when preceded by a neutral gender marking (e.g., *leur joli bateau*), a 19-msec difference. For the early bilinguals, the corresponding times were 481 and 525 msec, a 44-msec difference. The larger difference for the bilinguals appears to be due to their slightly longer mean reaction times in the neutral condition: 525 msec as compared with 498 msec for the monolinguals. Two-way ANOVAs confirmed these results. There was a strong congruency effect that was significant by participants and by items [ $F_1(1,30) = 60.16$ ,  $MS_e = 270.76$ ;  $F_2(1,35) = 48.54$ ,  $MS_e = 749.56$ ], a marginal group effect significant by items only [ $F_2(1,35) = 21.18$ ,  $MS_e = 364.94$ ], and a significant interaction [ $F_1(1,30) = 9.19$ ,  $MS_e = 270.76$ ;  $F_2(1,35) = 15.83$ ,  $MS_e = 354.56$ ]. A Tukey HSD post hoc test shows that both the monolinguals' 19-msec and the bilinguals' 44-msec congruency differences (neutral/correct) were significant at the .01 level. Thus we have replicated a congruency effect with monolinguals using naming, and we have shown that early bilinguals, like their monolingual counterparts, are sensitive to congruent gender marking. Will monolinguals also show an incongruency effect, and will early bilinguals be sensitive to incongruent marking? The answer can be observed in Figure 2.

Figure 2 presents the results of the incorrect/neutral part of the study. Mean naming times (in milliseconds) are plotted as a function of group (monolingual and early bilingual) and gender marking (neutral and incorrect). As can be seen, the incongruency effect was large for both groups: monolinguals took on average 483 msec to name items preceded by a neutral gender marking, but 513 msec when they were preceded by an incongruent gender marking, a difference of 30 msec. The corresponding values for the early bilinguals were 519 and 574 msec, a difference of 55 msec. This larger difference is in part due to the bilinguals' longer mean reaction times in the incongruent condition. The two ANOVAs showed that there was a strong incongruency effect significant by participants and by items [ $F_1(1,30) = 102.59$ ,  $MS_e = 280.06$ ;  $F_2(1,35) = 202.45$ ,  $MS_e = 320.15$ ], a marginal group effect by items only [ $F_2(1,35) = 240.22$ ,  $MS_e = 353.12$ ], and a significant interaction [ $F_1(1,30) = 10.03$ ,  $MS_e = 280.06$ ;  $F_2(1,35) = 20.46$ ,  $MS_e = 305.76$ ]. A Tukey HSD post hoc test shows that both the monolinguals' 30-msec and the bilinguals' 55-msec incongruency differences (incorrect – neutral) were significant at the .01 level.

We have therefore been able to replicate both a congruency (or facilitation) effect and an incongruency (or inhibition) effect with monolinguals in French by means of a naming task. To do this we divided up the three conditions

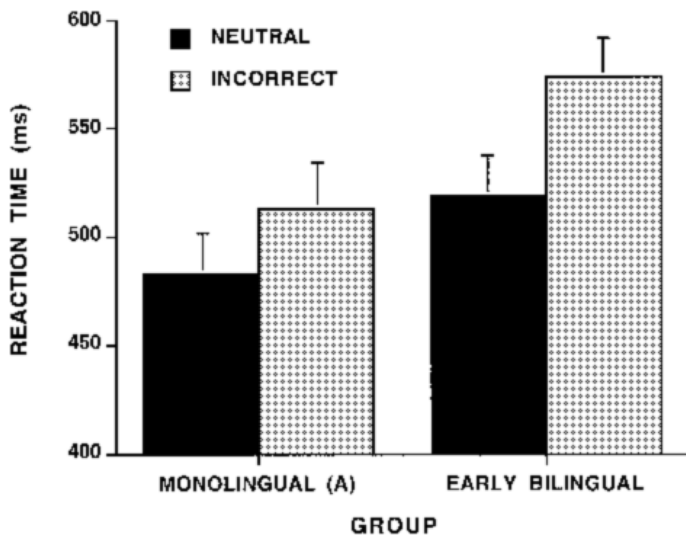


Figure 2. Mean naming times (in milliseconds) as a function of group (monolingual [A] and early bilingual) and gender marking (neutral and incorrect). Error bars represent  $+1$  SEM.

(correct, neutral, and incorrect) into two groups of two (correct and neutral, incorrect and neutral). In addition, we have shown that bilinguals who acquired and started using a gender-agreement language, along with another language at age 5;4 years on average, demonstrated strong congruency and incongruency effects. They have become sensitive to gender early in life and they appear to use gender marking in perception the way monolinguals do. The only apparent difference with monolinguals is that they appear to be even more sensitive to gender congruency and incongruency, as can be seen by the larger differences between the neutral and the other two conditions. It is unclear why this occurred, and only future studies will be able to assess its importance.

The crucial question now becomes: Will late bilinguals show the same effects as early bilinguals? If gender marking is indeed important during language processing, then they should become sensitive to it. However, if there is a "critical period" for taking into account gender marking (in perception at least), and if late bilinguals acquired, and started using on a regular basis, their gender-marking language after this period, then they should show little if any effect.

## EXPERIMENT 2 Late Bilinguals

In this experiment, we examined the behavior of late English–French bilinguals and we compared it with that of French monolinguals. Since the late bilinguals were in their late 40s on average, we made sure that the monolinguals were of a comparable age.

### Method

**Participants.** Two groups of participants were used in this study. The first was made up of 32 French monolinguals (randomly as-

signed to two groups of 16) who were matched on age with the late bilinguals (46;6 and 48;5 years, respectively). The monolinguals reported no speech or hearing defects, and they served individually in the experiment. The second group of participants was made up of 32 late English–French bilinguals (assigned randomly to one of two groups of 16) who reported no speech or hearing defects and who also served individually in the study. Their first language was English and they had started learning French in school. They became regular users of both languages only when they moved to a French-speaking country as adults (the mean age of onset of their bilingualism was 24;8). Table 1 (right columns) presents the group's biographical and language proficiency data. We note that the late bilinguals reported starting to speak English at age 1;0 year, on average, and French at age 15.11 (the early bilinguals had means of 1;11 and 4;0, respectively). The large gap between starting to speak French and the onset of bilingualism (some 9 years later) is due to the fact that at first French was simply a language being learned at school and not a language of everyday interaction. It started being so only when they moved to a French-speaking country. In regard to reported language proficiency, the late bilinguals reported near-perfect English oral comprehension (mean of 6.8) and a slightly lower level of French comprehension (6.0). Their reported level of oral language production favored English (6.7) over French (5.4). It is clear that the late bilinguals were dominant in English, but what is important for our purpose is that they rated their oral comprehension in French as being very good (6.0 on a 7-point scale). This is not surprising since they have been active bilinguals for up to 24 years on average (their mean age is 48;5 years).

**Materials and Procedure.** Exactly the same materials and procedure were used as in the first experiment. Once again, French was used throughout the testing session, which was conducted by a native speaker of French (D.G.).

### Results and Discussion

The data obtained from the two groups (late bilinguals and monolingual controls) were analyzed in the same way as the data in first study. For the late bilinguals, in the correct/neutral part of the study, 1.56% of the times were greater than 1,000 msec, 3.65% of the values were above or below  $2$  SD, and 3.13% of the values were missing. In the incorrect/neutral part, 1.74% of the times were greater than 1,000 msec, 4.17% of the values were above or below  $2$  SD, and 1.56% of the values were missing. For the monolingual controls, in the correct/neutral part of the study, 0.17% of the times were greater than 1,000 msec, 3.99% of the values were above or below  $2$  SD, and 2.60% of the values were missing. In the incorrect/neutral part, 0.17% of the times were greater than 1,000 msec, 2.60% of the values were above or below  $2$  SD, and 2.43% of the values were missing. The data replacement procedures were the same as in the first experiment.

Figure 3 presents mean naming times (in milliseconds) as a function of group—the monolingual control group, labeled "Monolingual (B)", and the late bilingual group—and gender marking (correct and neutral). First, we can see that the monolingual group showed a congruency effect: Mean naming times for the correct and the neutral conditions were 521 and 545 msec, respectively (a 24-msec congruency difference). Second, we note that the late bilinguals had longer reaction times than the controls, but we are struck especially by the fact that they showed ab-

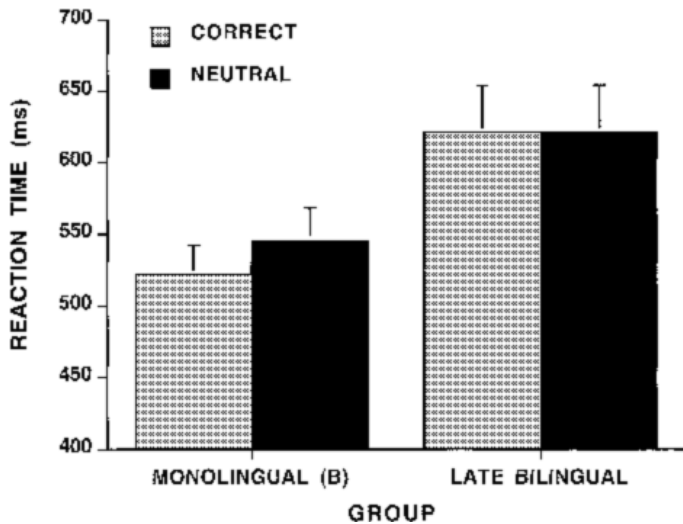


Figure 3. Mean naming times (in milliseconds) as a function of group (monolingual [B] and late bilingual) and gender marking (correct and neutral). Error bars represent +1 SEM.

solutely no naming time difference between the correct and neutral conditions: Both took 620 msec on average. The two-way ANOVAs confirm these results. There was a congruency effect [ $F_1(1,30) = 20.89, MS_e = 120.88; F_2(1,35) = 10.51, MS_e = 542.35$ ], a group effect [ $F_1(1,30) = 4.94, MS_e = 24,603.92; F_2(1,35) = 375.02, MS_e = 729.83$ ], and an interaction [ $F_1(1,30) = 19.66, MS_e = 120.88; F_2(1,35) = 9.00, MS_e = 594.78$ ]. A Tukey HSD post hoc test shows that the monolinguals' 24-msec congruency difference (neutral – correct) was significant at the .01 level. Although the late bilinguals showed no congruency effect, we can wonder whether they might not be sensitive to gender incongruity. After all, it is a phenomenon that often reaches consciousness, as can be seen when native speakers react sometimes quite strongly to gender-production errors (see Grosjean et al., 1994). The answer is in Figure 4.

Figure 4 presents mean naming times (in milliseconds) as a function of group (monolingual and late bilingual) and gender marking (neutral and incorrect). As can be seen, the monolinguals showed a large incongruency effect: Their mean naming time was 547 msec in the neutral condition and 594 msec in the incorrect condition, a difference of 47 msec. The late bilinguals responded a bit more slowly than the monolinguals, but above all they showed no apparent difference between the two conditions: Their mean naming times were 632 msec in the neutral condition and 626 msec in the incorrect condition, a 6-msec difference in the opposite direction! An ANOVA shows a marginal incongruency effect by participants only [ $F_1(1,30) = 43.09, MS_e = 161.83$ ], a marginal group effect by items only [ $F_2(1,35) = 131.44, MS_e = 937.29$ ], but a strong interaction [ $F_1(1,30) = 68.13, MS_e = 161.83; F_2(1,35) = 22.32, MS_e = 1,108.95$ ]. A Tukey HSD post hoc test shows that the monolinguals' 47-msec incongruency difference (incorrect – neutral) was significant at the .01 level, whereas the bilinguals' 6-msec difference in the opposite direction was not significant.

It is clear from both these sets of results that late bilinguals are insensitive to both gender congruency and gender incongruency. It is as if they just cannot use the masculine *le* cue or the feminine *la* cue during the processing of the noun phrase. In order to make sure that this absence of an effect is not simply due to overall speed of response (the bilinguals were somewhat slower than the controls, and this might have neutralized the gender-marking effect), we took the participants' mean reaction times to the neutral stimuli, in each condition and in each group, and used them to divide the participants into two subgroups: the slowest 8 and the fastest 8. We then obtained subgroup means and tested them for a difference. Whereas the monolinguals showed both a congruency and an incongruency effect, whatever subgroup they belonged to, the bilinguals showed no effect, even though the bilinguals in the fast subgroups were faster than the monolinguals in the slow subgroups. We concluded that overall speed of response cannot account for the late bilinguals' absence of a gender-marking effect.

A second question we asked was whether the total lack of a congruency and incongruency effect could be linked, to some extent at least, to the late bilinguals' inability to use gender agreement when speaking French. Could the problem be linked to production, at least in part? The fact that the late bilinguals cannot produce the correct gender when speaking might explain why they could not use gender marking during perception. In order to assess this, 1 month after finishing the second experiment, we phoned 15 participants from each of the two bilinguals groups, late bilinguals and early bilinguals, and asked them to repeat back each of the 36 nouns used in the studies but to precede them with the appropriate determiner *le* or *la*. Thus, for example, if we gave them *bateau*, they were to say *le bateau*. (A few practice examples at the start resolved any problems they may have had understanding the task.) As expected, the early bilinguals did not make a single gender-production error (a mean of 0 errors on 36 possible errors), and the late bilinguals made only very few errors (a mean of 3.5 errors on 36 possible errors, with a standard deviation of 3.2). With a bit less than 10% errors, therefore, one can conclude that the late bilinguals know the gender of French nouns (at least that of common nouns such as those used in the experiment), and they can produce the appropriate determiner when asked to. The processing problem they had in perception, therefore, is not linked to one in production, at least not directly.

Two anonymous reviewers proposed factors related to language proficiency to explain the lack of any effects in the late bilinguals. The first factor relates to the discriminability of the three determiners, *le*, *la*, and *leur*, by these participants. The argument is that they simply cannot hear the difference and hence will not react any differently to the three conditions. A number of counterarguments can be proposed, however. First, great care was taken to use quality exemplars of these determiners (see the Materials section of the first experiment). Second,

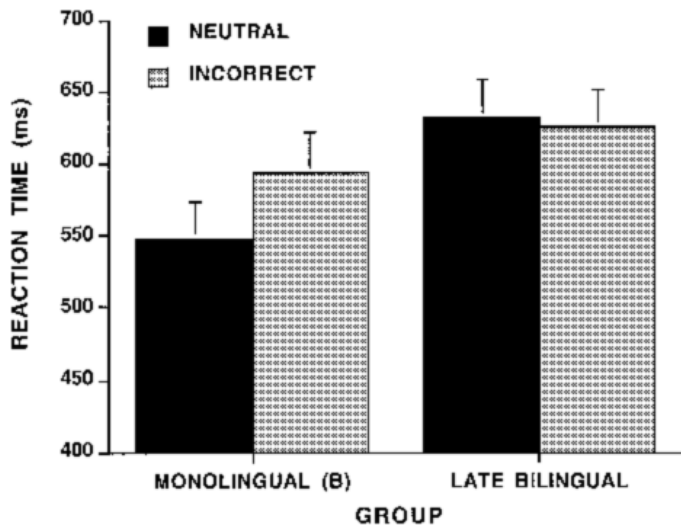


Figure 4. Mean naming times (in milliseconds) as a function of group (monolingual [B] and late bilingual) and gender marking (neutral and incorrect). Error bars represent +1 SEM.

the phonetic difference is relatively large between the three: a central vowel for the *le*, a low front vowel for the *la*, and the presence of an additional consonant (/R/) for the *leur*. Finally, it should be recalled that the late bilinguals had been active bilinguals for up to 24 years, listening and speaking French on a daily basis, and hence were used to discriminating vowel sounds. The second factor put forward is that of fluency. It was proposed that fluency is driving the effect, so that the more fluent a person is in the gender-marking language, the more he/she will be sensitive to gender marking. Although further research is needed to fully assess the role of this factor, one should recall that the late bilinguals rated their oral comprehension of French as very high (a mean of 6.0 on a 7-point scale). Admittedly, no objective measure of fluency was obtained, but there is no reason to believe that the late bilinguals' oral comprehension of French was not good (or that their self-ratings were erroneous). In addition to their residency of more than 20 years in a French-speaking country, many have spouses and children with whom they speak French on a daily basis.

We can conclude from the above that speed of response, production skills, and language proficiency are not clear candidates to account for the absence of a gender-marking effect in late bilinguals. Since age does not seem to be a problem either (it was controlled for by testing monolinguals of a similar age), we must conclude that the absence of the gender-marking effects is probably due to when the gender-agreement language was acquired and started to be used on a regular basis.

## GENERAL DISCUSSION

In this paper, we replicated in French, and by means of a naming task, the gender congruency and incongruency effects found by Bates et al. (1996) in Italian.<sup>4</sup> We then investigated whether bilinguals process gender marking,

and we found that this depends on when they acquired the gender-agreement language and started using it on a regular basis. We first showed that early English–French bilinguals behave like monolinguals both on gender congruency and gender incongruency. The only possible difference between the two groups could be that the early bilinguals may have been more sensitive to gender marking. When we tested late bilinguals, however, we found a total insensitivity to gender marking in perception, whether the gender was congruent or incongruent, although a control group of monolinguals of the same age showed both effects. An analysis in terms of speed of response and a small follow-up production study allowed us to exclude the late bilinguals' speed of response and their gender-production skills as factors accounting for the absence of the gender-marking effect.

As we indicated in the beginning of this paper, there are two possible accounts of the gender-marking effect. On the one hand, the process could be internal to the lexicon in that when a word carrying a gender marking is identified, it activates all the nouns in the lexicon that share the same gender. A variant of this first possibility is that it is the gender feature of the word that activates all the nouns with that same feature. To account for the results obtained with late bilinguals, a first suggestion could be that they have not established any gender connections among the words sharing the same gender or that they have not given a gender feature to the nouns. However, since they did extremely well on the follow-up production task, one would probably have to conclude that they do have these connections (or features), but that they simply do not activate them during auditory processing. The alternative explanation for the gender-marking effect involves both the lexical and syntactic processing modules. The lexical module undertakes the recognition of the words in question (e.g., an article and a noun), and the syntactic module checks that gender agreement is respected. In the case of late bilinguals who do not show a gender-marking effect, either they never developed (or "triggered") this mechanism or they simply cannot make use of it (in perception at least).

Whatever the account, it would seem that certain processing mechanisms in a second language will never be acquired (or only partly acquired) after a specific point. In the case of gender, Carroll (1989) has proposed a detailed account of how first- and second-language learners acquire gender. According to her, when first-language learners (which in our case would correspond to our monolinguals and our early bilinguals) figure out that determiners are distinct lexical items, the phonological representations are reduced and the morphosyntactic representations are augmented to include the feature [ $\pm$  masc]. Thus, the reanalysis of determiners as separate words serves as a trigger for the activation of the universally specified gender feature. If the gender feature is not needed, it atrophies and disappears. In the case of second-language learners (our late bilinguals in this case), Carroll states that the functions deriving underlying phono-

logical representations will not chunk determiners with nouns and hence will not include a gender feature. To produce gender, second-language learners must develop mnemonic strategies for pairing nouns and all gender-marked words; to do so they develop rules of thumb that correspond to preference rule systems. This might explain why our late bilinguals managed to produce 90% of the appropriate determiners when asked to say the nouns preceded by the appropriate definite article, *le* or *la*. Carroll concludes that anglophones will have no difficulty "hearing" the words that mark gender when parsing speech because they can phonologically represent all forms and carry out lexical look-up. On this point, if hearing means processing, then Carroll's statement needs to be modulated somewhat. Words will indeed be recognized (our late bilinguals reported having very good French oral comprehension), but lexical access will not be speeded up by a congruent gender marking on the preceding word(s) or slowed down by an incongruent gender marking. In other words, late bilinguals cannot call on gender marking to facilitate (or, more rarely, impede) the word recognition process.

The present study raises many interesting questions that will need to be addressed in future work. First, it would be interesting to see whether late bilinguals recognize correct and wrong gender (using grammaticality judgments, for example) even though they are not sensitive to the phenomenon during perception. They might well do so as different, more metalinguistic, skills are required. Second, it will be important to use different word recognition tasks to see if the difference between early and late bilinguals is maintained. Even though auditory naming is not a "shallow" task (see the various effects found with the task, some quite "deep"; Bates & Liu, 1997), it could be that the type of task used plays a role in some way when testing late bilinguals (hopefully a very weak one). Third, and linked to this last point, it will be important to test early bilinguals who are middle aged since it is always difficult to compare participants across age groups (a suggestion made by one of the reviewers). It should be recalled that the early bilinguals were some 24 years younger than the late bilinguals. Fourth, concerning the relationship between gender production and perception, one could investigate how late bilinguals behave on a speeded production task. They would probably make many more errors than in our follow-up study (we put no pressure on them to respond quickly), but they would probably not reach chance level (i.e., 50% errors, which would be the production equivalent of the absence of a gender-marking effect in perception). If they didn't reach this error level, we would have to conclude that gender perception and production are governed by somewhat different processing mechanisms. Fifth, we should study whether late bilinguals are more sensitive to gender markings if there are more of them. In our study, only the determiner carried gender information (the following adjective *joli(e)* did not), and it might have been insuffi-

cient information for our late bilinguals. There are many instances in French where several gender markings precede the noun, such as *la belle journée* (*the nice day*), where both *la* and *belle* carry a feminine gender marking, and late bilinguals might just become sensitive to gender marking when there are more gender cues. Finally, we can ask whether the problem we have uncovered with our late bilinguals is a general problem of agreement or whether it is limited to gender. What would happen, for example, with number agreement such as in *les beaux arbres* (*the nice trees*) where *les* and *beaux* both carry a plural number marking? Of course, one would have to test pairs of languages where one language does not have overt number agreement (both English and French do), but it would nevertheless be interesting to see whether the processing pattern is the same as with gender.

In sum, bilinguals do indeed use gender marking during spoken word recognition, but only if they started acquiring, and using on a regular basis, the gender-agreement language at an early age. Late bilinguals do not seem to be able to use gender marking during the recognition process. One can therefore extend Sir Winston Churchill's statement in the following way "I have never mastered the gender of French nouns . . . *be it in production OR perception.*"

## REFERENCES

- BATES, E., DEVESCOVI, A., HERNANDEZ, A., & PIZZAMIGLIO, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*, 992-1004.
- BATES, E., DEVESCOVI, A., PIZZAMIGLIO, L., D'AMICO, S., & HERNANDEZ, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, *57*, 847-862.
- BATES, E., & LIU, H. (1997). Cued shadowing. In F. Grosjean, & U. Frauenfelder (Eds.), *A guide to spoken word recognition paradigms* (pp. 577-583). Hove, U.K.: Psychology Press.
- CACCIARI, C., CARREIRAS, M., & BARBOLINI CIONINI, C. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory & Language*, *37*, 517-532.
- CARROLL, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, *39*, 535-594.
- COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271.
- COLÉ, P., & SEGUI, J. (1994). Grammatical incongruity and vocabulary types. *Memory & Cognition*, *22*, 387-394.
- CONTENT, A., MOUSTY, P., & RADEAU, M. (1990). BRULEX: Une base de données lexicales informatisées pour le français écrit et parlé. *L'Année Psychologique*, *90*, 551-566.
- CORBETT, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- GOLLAN, T., & FROST, R. (2001). *Two routes to gender: Evidence from Hebrew*. Manuscript submitted for publication.
- GROSJEAN, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1-22). Oxford: Blackwell.
- GROSJEAN, F., DOMMERGUES, J.-Y., CORNU, E., GUILLELMON, D., & BESSON, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics*, *56*, 590-598.
- GURJANOV, M., LUKATELA, G., LUKATELA, K., SAVIC, M., & TURVEY, M. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 692-701.



JAKUBOWICZ, C., & FAUSSART, C. (1998). Gender agreement in the processing of spoken French. *Journal of Psycholinguistic Research*, *27*, 597-617.

PEARLMUTTER, N., GARNSEY, S., & BOCK, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory & Language*, *41*, 427-456.

*Le Robert Oral-Ecrit*. (1989). Paris: Les dictionnaires Le Robert.

ROGERS, M. (1987). Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics*, *8*, 48-74.

SCHMIDT, R. (1986). Was weiss der Artikel vom Hauptwort? Ein Beitrag zur Verarbeitung syntaktischer Beziehungen beim Lesen [What does the article know about the noun? A study of syntactic agreement processing during reading.]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *23*, 150-163.

SOARES, C., & GROSJEAN, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition*, *12*, 380-386.

VAN BERKUM, J. (1996). *The psycholinguistics of grammatical gender*. Nijmegen: Nijmegen University Press.

*Webster's Ninth New Collegiate Dictionary*. (1991). Springfield, MA: Merriam-Webster.

NOTES

1. It should be recalled that the lexical processing module is used to recognize words and to ensure that all the information needed for further processing is extracted from the mental lexicon, whereas the syntactic processing module identifies and structures the constituents.
2. The reasons that led us to choose naming over lexical decision are linked to the bilinguals. First, some bilinguals (especially late bilinguals) do not always feel secure deciding whether an item is a word or a nonword in their second language, and this probably has an impact on their processing. Second, reaction times to nonwords are longer in bilinguals than in monolinguals, as shown by Soares and Grosjean (1984). They explained this finding by suggesting that bilinguals search both lexicons when confronted with a nonword. We wanted the other language (English) to be as inactive as possible during the experiment (at least not above a residual activation level that is probably always there; Grosjean, 2001), and hence we opted not to use lexical decision in this study.
3. When this experiment was done, some doubt existed as to the reliability of the voice-operated relay in the PsyScope button box. We

therefore decided to use an external counter that we knew from past studies to be extremely reliable.

4. In an unpublished study, we found similar congruency and incongruency effects in French with a lexical decision task. The stimuli were the same as those of this paper. The results are available upon request.

**APPENDIX**  
**The 36 Words Used in the Study Accompanied**  
**by the Translation of Their Most Frequent Meanings**

Masculine Nouns	
One Syllable	Two Syllables
camp (camp)	bateau (boat)
plat (dish)	plateau (tray)
drap (sheet)	poisson (fish)
teint (complexion)	cadeau (present)
pré (meadow)	bassin (pool)
puits (well)	bijou (jewel)
disque (record)	berceau (cradle)
clan (clan)	pinceau (brush)
tube (tube)	dessert (desert)
Feminine Nouns	
One Syllable	Two Syllables
glace (ice)	bouteille (bottle)
plante (plant)	boutique (store)
danse (dance)	galerie (gallery)
cave (cellar)	pension (pension)
poule (hen)	barrière (fence)
cage (cage)	bougie (candle)
boucle (buckle)	balance (scales)
torche (flashlight)	trompette (trumpet)
pelle (shovel)	poupée (doll)

(Manuscript received February 7, 2000;  
revision accepted for publication August 15, 2000.)