Cognition 114 (2010) 266–275

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Limits on bilingualism revisited: Stress 'deafness' in simultaneous French–Spanish bilinguals

Emmanuel Dupoux[a,*], Sharon Peperkamp[a,b], Núria Sebastián-Gallés[c]

[a] Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole des Hautes Etudes en Sciences Sociales, Département d'Etudes Cognitives, Ecole Normale Supérieure, Centre National de la Recherche Scientifique, 29 Rue d'Ulm, 75005 Paris, France
[b] Département des Sciences du Langage, Université de Paris VIII, Saint-Denis, France
[c] Brain and Cognition Unit, Universitat Pompeu Fabra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 29 September 2008
Revised 30 September 2009
Accepted 1 October 2009

Keywords:
Speech perception
Simultaneous bilingualism
Phonology
Stress

ABSTRACT

We probed simultaneous French–Spanish bilinguals for the perception of Spanish lexical stress using three tasks, two short-term memory encoding tasks and a speeded lexical decision. In all three tasks, the performance of the group of simultaneous bilinguals was intermediate between that of native speakers of Spanish on the one hand and French late learners of Spanish on the other hand. Using a composite stress 'deafness' index measure computed over the results of the three tasks, we found that the performance of the simultaneous bilinguals is best fitted by a bimodal distribution that corresponds to a mixture of the performance distributions of the two control groups. Correlation analyses showed that the variables explaining language dominance are linked to early language exposure. These findings are discussed in light of theories of language processing in bilinguals.

1. Introduction

How many languages can be sustained efficiently in a single brain? Much research has been devoted to the acquisition of a second language by monolingual individuals, and in particular to the negative correlation between the age of acquisition and the outcome of second language learning: Late learners typically have worse performance in their second language than early learners (Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989; Weber-Fox & Neville, 2001), even though considerable individual variation exists and it has been argued that native-like performance can be attained by some individual late learners (Birdsong, 2007; Bongaerts, 1999). Little research, in contrast, has been devoted to the study of *simultaneous bilinguals*, that is, individuals who are immersed from birth in a bilingual environment. Given the above-mentioned findings with late learners, the expectation is that given a relatively balanced exposure to both languages, simultaneous bilinguals should easily attain native-like performance in both languages. Yet two studies, concerned with French–English and Spanish–Catalan bilinguals, respectively, claimed that simultaneous bilinguals may not reach native-like performance in one of their languages (Cutler, Mehler, Norris, & Segui, 1989, 1992; Sebastián-Gallés, Echeverria, & Bosch, 2005). Both showed that the performance of even extremely proficient bilinguals who pass for perfect monolinguals in both of their languages in ordinary language situations may be below native levels in at least one of the languages when tested with sophisticated laboratory techniques.

Cutler et al. (1989, 1992) raised the possibility that a bilingual brain processes only one language with optimal efficiency; the other language is processed using routines of the primary language, hence yielding non-native performance. Given a group of simultaneous bilinguals, this hypothesis predicts that for each of the two languages performance does not follow a normal, monomodal, distribution, but rather a bimodal one, with participants in the best

* Corresponding author. Tel.: +33 1 44 32 26 17.
E-mail address: emmanuel.dupoux@gmail.com (E. Dupoux).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.10.001

mode performing as monolinguals, and participants in the other mode performing as second language learners. This prediction is not strongly supported by empirical data: Both Cutler et al. (1989, 1992) and Sebastián-Gallés et al. (2005) used several measures of *dominance*, defined through a subjective questionnaire or a biographical interview, to divide their bilingual participants into two groups. At least one of these measures yielded groups with performance differences. However, this does not imply that the bilinguals' performance followed a bimodal distribution. Indeed, the same difference could result from a monomodal distribution with a large variance, hence with only few (if any) bilinguals performing like monolinguals of one or the other language. Moreover, whereas Cutler et al. (1989, 1992) found that – as predicted – the performance of the best group was like that of monolingual speakers and that of the other group like that of second language learners[1], Sebastián-Gallés et al. (2005) reported contrasting results. They found, firstly, that the performance of the best group was *worse* than that of native speakers, suggesting that the performance in the dominant language is influenced by the other language, akin to what has been found for speech production in both early and late bilinguals (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Flege, Schirru, & MacKay, 2003; Yeni-Komshian, Flege, & Liu, 2000). Secondly, they found that the performance of the other group was *better* than those of early Spanish–Catalan bilinguals, i.e. native speakers of Spanish who had started to learn Catalan around the age of four. Taken together, their results thus suggest that if the performance of simultaneous bilinguals is bimodal (which remains an open question), then the two modes are different from those defined by the performance of monolinguals on the one hand and that of second language learners on the other hand.

The hypothesis that a bilingual brain processes only one language with optimal efficiency also raises a question concerning acquisition: which factor(s) determine(s) which of the languages spoken in the infant's environment will become his or her dominant language? No agreement emerges from the previous studies with respect to this question. The bilinguals tested by Cutler et al. could be partitioned into French-dominant versus English-dominant groups on the basis of their preferred language, whereas in the study by Sebastián-Gallés et al., the most relevant criterion for partitioning the bilinguals into a Catalan-dominant and a Spanish-dominant group was the language of the mother. Resolving this issue is important for theories of early language development and functional brain plasticity. Indeed, theories based on an early attunement of speech processes predict an effect of early exposure (Kuhl, 2000); theories based on the existence of a sensitive period (Weber-Fox & Neville, 2001) predict no (or little) effect of exposure that follows the sensitive period; and theories based on life-long plasticity predict either an effect of total exposure, or an effect of the later occurring exposure (Birdsong, 2005; Hakuta, Bialystok, & Wiley, 2003).

[1] Note, though, that no statistical analyses are shown to back up this finding.

In the present study, we consider a group of French–Spanish simultaneous bilinguals and focus on their performance in one of their languages, i.e. French. Specifically, we test their on-line perception of word-level stress, a phonological feature that is used in Spanish but not French.[2] In earlier work, it was demonstrated that native speakers of French, as opposed to native speakers of Spanish, exhibit a robust stress 'deafness', that is, they have much difficulty perceiving stress contrasts (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001), even after having learned Spanish as a second language (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). The strength of the stress 'deafness' effect places us in a good starting position to analyze the underlying distribution of the simultaneous bilinguals' performance. In addition, we increase the statistical power of our data by calculating a composite score based on several tasks. The reasoning is as follows: any given paradigm is loaded with task-specific components (memory, attention, executive functions, metalinguistic skills), which are orthogonal to the particular processing level under study. Individual variation in these irrelevant components is bound to blur the shape of the distribution, making it difficult to distinguish between a monomodal and a bimodal distribution. Taking into account the results across distinct tasks should allow us to reduce the contribution of these task-specific components, but retain the contribution of the processing levels that are common to the tasks. This power will also enable us to test whether the two modes are similar to the modes of monolingual or non-native control populations.

In order to address the question as to which factor(s) determine(s) which of the languages spoken in the infant's environment will become his or her dominant language, we selected our simultaneous bilinguals in two different cities, one French and one Spanish; we conducted individual biographic interviews with each one of the bilinguals; we had them fill in a questionnaire concerning a subjective assessment of their language competences and preferences; and we had their parents fill in an extensive questionnaire concerning their early language exposure. We used these data to quantify the amounts of exposure to French and Spanish throughout the bilinguals' lives (prenatal influences, infancy, childhood, adolescence, adulthood), and to establish their fluency in French and Spanish as well as the importance they attach to their two languages.

2. Sample

Twenty-three simultaneous French–Spanish bilinguals participated. There were five men and 18 women, aged between 19 and 44 (mean: 25). Ten bilinguals lived and were tested in Paris, 12 lived and were tested in Barcelona, and one lived in Barcelona but was tested during a short visit in Paris. Fifteen bilinguals had a native French-speaking mother and a native Spanish-speaking father, while the

[2] A pilot study with eight simultaneous French–Spanish bilinguals was reported in Peperkamp, Dupoux, and Sebastián-Gallés (1999). In this pilot, three bilinguals showed French-like results and the remaining five Spanish-like results, with individual performance being highly correlated with the country of birth.

Table 1
Language background and subjective language assessment of 23 simultaneous bilingual participants. SDs are shown between parentheses.

Language background	
Percentage of Spanish mothers	35%
Percentage of early (0–2 years) exposure to Spanish	38%
Estimated by participants' parents	38%
Inferred from participants' interview	41%
Percentage of participants resident in a Spanish-speaking country[a]	
0–2 years	65%
2–4 years	63%
4–10 years	54%
10–18 years	54%
At time of test	57%
Percentage of use of Spanish at time of test	56%
Subjective language assessment	
Importance (*on a scale from* 1 *to* 10)	
Spanish	9.17 (0.78)
French	9.52 (0.73)
Pronunciation (*on a scale from* 1 *to* 10)	
Spanish	8.96 (1.22)
French	9.17 (1.34)
Grammar (*on a scale from* 1 *to* 10)	
Spanish	8.67 (1.06)
French	8.78 (1.65)
Vocabulary (*on a scale from* 1 *to* 10)	
Spanish	8.33 (1.04)
French	8.65 (1.34)

[a] Participants who moved from one country to another during the reference period were counted on a pro rata basis.

remaining eight had a native Spanish-speaking mother and a native French-speaking father.

There were two control samples. The first one consisted of 20 native speakers of Spanish, the same ones that constituted the Spanish control sample in Dupoux et al. (2008). There were three men and 17 women, aged between 18 and 25 (mean: 21). All had been born in monolingual Spanish families in Barcelona, the city in which they were tested.[3] The second control sample consisted of the 39 native speakers of French who had learned Spanish after age 10, tested by Dupoux et al. (2008).[4] There were 11 men and 28 women, aged between 20 and 57 (mean: 28). Six of them lived and were tested in Paris, the remaining 33 lived and were tested in Barcelona.

None of the participants had a known hearing deficit.

2.1. Language background

The 23 simultaneous bilinguals were interviewed concerning their language background (see Table 1). For five periods of their lives (0–2 years, 2–4 years, 4–10 years, 10–18 years, and 18-present), they indicated where they had lived (some of them had moved at least once to another country, which is not surprising given their mixed-language family situation). They also provided more detailed information concerning their first two years of life (who took care of them and in what language they were spoken to by various people in their environment) as well as their current situation (language(s) spoken at home, at school/work, with friends and siblings; frequency and length of visits to Spanish-speaking countries for those living in France and to French-speaking countries for those living in Spain). On the basis of this information (and without knowing the participants' scores on the stress perception tasks), the three authors independently estimated for each bilingual his/her amount of exposure to Spanish during the first two years of life (three categories: 0–40%, 40–60%, and 60–100%), as well as his/her current use of Spanish (on a scale from 1 to 5). They then compared their classifications and discussed cases of disagreement. If the disagreement persisted, the final classification was the one that had been proposed by two out of three authors (there were no cases in which each author had proposed a different classification).

In addition, the parents of all but one of the bilinguals filled in a detailed questionnaire concerning the first two years of their child's life, culminating in an estimation of the percentages of exposure to French and Spanish (see Table 1). This questionnaire asked in particular which languages were spoken to the child by the mother, the father, and any other family members or people living with the family (such as siblings, grandparents, or a nanny), how much time each of these people spent with the child, and whether or not the child went to a daycare center (and if so, for how many hours a week).

Finally, the bilinguals rated on a scale from 1 to 10 the importance of French and Spanish to their own lives, as well as their current French and Spanish competence in three domains, i.e. pronunciation, grammar, and vocabulary (see Table 1).

3. Experiment

We used three paradigms that have been shown to be very sensitive to the phonological encoding of lexical stress during on-line perception. The first two used two different versions of a sequence recall task in which participants have to quickly and efficiently recode sequences of speech sounds into a short-term memory phonological store. Specifically, they first learn to associate the members of a minimal pair to the keys [1] and [2] of a computer keyboard, and then listen to sequences constituted by repetitions of the two non-words which they have to reproduce by typing the associated keys in the correct order. Native speakers of French, contrary to native speakers of Spanish, have much difficulty performing this task on sequences involving a stress contrast. In particular, the version we used in the first task has been shown to separate French and Spanish monolinguals from one another at the level of individual participants; that is, the performance distributions of the French and the Spanish are completely non-overlapping (Dupoux et al., 2001). The version we used in the second task has been shown to separate French late learners of Spanish from Spanish monolinguals, with the performance of the former being significantly different from that of the latter (Dupoux et al., 2008). Finally, in the third task, we use speeded lexical decision with words and non-words that differ only in lexical stress; this task has simi-

[3] Barcelona being a bilingual Catalan–Spanish city, these participants were also fluent in Catalan, whose stress pattern is very similar to the Spanish one.

[4] Given that one of the tasks we use is lexical decision, the control group had to consist of native speakers of French with knowledge of Spanish.

larly been shown to separate French late learners of Spanish from Spanish monolinguals (Dupoux et al., 2008).

We analyze these three tasks in the following way. First, we perform group analyses, comparing the simultaneous bilinguals with the Spanish and French controls on the three tasks separately. Next, we compute an individual composite score over the three tasks and examine the distribution of the simultaneous bilinguals' scores to those of the Spanish and French controls, respectively. Finally, we run correlation analyses on the composite score against a set of biographical variables.

3.1. Method

3.1.1. Task 1: stimuli and procedure

The first sequence recall task used the same stimuli and procedure as those of Experiment 3 in Dupoux et al. (2001).

Two minimal pairs of non-words (both in French and in Spanish) were used, one with a phonemic contrast (/kúpi/–/kúti/), the other with a stress contrast (/mípa/–/mipá/). For both contrasts, the first non-word was associated with key [1] and the second one with key [2] of computer keyboard. Each item was instantiated by six tokens, produced by a female speaker. As to the tokens with the stress contrast, stressed vowels were longer than unstressed vowels, they had a higher F0, and they were louder (for details, see Dupoux et al., 2001).

Participants were first tested on the phoneme contrast.[5] They were told that they were going to learn two words in a foreign language, and were first asked to press the number key [1], upon which they heard all tokens of the first item. They were then asked to press the number key [2], upon which they heard all tokens of the second item. Subsequently, they could continue listening to the various tokens of the two items by pressing the associated keys; pressing each one of these keys resulted in the playing of one token of the corresponding item. Participants could thus hear as many tokens of the two items as they wished. When they indicated they were ready to move on, they were trained on the distinction between the two items as well as the correct association between the items and the number keys. They heard a token of one of the items and had to press the associated key, [1] or [2]. A message on the screen informed participants whether their responses were correct. We defined a success criterion of seven correct responses in a row. After having reached this criterion, participants turned to the main experiment, consisting in the oral presentation of a sequence constituted by repetitions of the two items, with an ISI of 80 ms. There were 40 trials, presented in five blocks of eight. The trials in the different blocks contained sequences of length 2, 3, 4, 5, and 6, respectively.

The participants' task was to reproduce each sequence by typing the associated keys in the correct order. For instance, the sequence of length 4 /kuti–kuti–kupi–kuti/ was to be transcribed as 2212. For each participant, the order of sequences was randomized, and each item was instantiated randomly by a token with the proviso that a single token could not appear more than once in a sequence. Each trial consisted of a sequence followed by the word 'OK', and participants could not begin typing their response until they had heard this word. Participants were warned whenever they entered a sequence with a length that did not correspond to the length of the input string and asked to enter their reply again. A 1500 ms pause separated each response from the next trial.

After the last test trial, the whole procedure was repeated with the minimal pair containing the stress contrast.

Responses that are a 100% correct transcription of the input sequence were coded as correct; all other responses as incorrect. Among the incorrect responses, those that are a 100% incorrect transcription – i.e. with each token of the sequence labeled incorrectly – were coded as reversals. Participants with more reversals than correct responses in either the phoneme or the stress condition were rejected, the high percentage of reversals suggesting that they might have confused the number key associated to the first item with the one associated to the second item.

3.1.2. Task 2: stimuli and procedure

The second sequence recall task used the same stimuli and procedure as those of Experiment 1 in Dupoux et al. (2008).

As in the previous task, two minimal pairs of non-words were used, one with a phonemic contrast (/fíku/–/fítu/), the other with a stress contrast (/númi/–/numí/). Each item was instantiated by six tokens produced by six different speakers, three male and three female. As to the tokens with the stress contrast, stressed vowels were longer than unstressed vowels, they had a higher F0, and they were louder (for details, see Dupoux et al., 2008).

The procedure was identical to the one in the first task, with two exceptions. First, the ISI was shortened from 80 to 50 ms. Second, the main experiment consisted of four warm-up trials with sequences of length 2 and 28 test trials with sequences of length four. During the warm-up phase, incorrect responses gave rise to an error message that was displayed for 800 ms, after which the same sequence was repeated until the correct response was given. During the test phase, no feedback was provided and no sequences were repeated.

3.1.3. Task 3: stimuli and procedure

The lexical decision task used the same stimuli and procedure as the one in Dupoux et al. (2008).

The stimuli consisted of pairs of words and non-words, recorded by a female Spanish native speaker (for a complete list, see Appendix B of Dupoux et al. (2008)). There were 112 experimental pairs that differed only in the position of stress, used for the stress condition. Ninety-six of them were disyllabic, half of the real words having initial stress and the other half final stress; the remaining 16 items were trisyllabic (2 words with initial stress, 6 with medial stress and 8 with final stress). An additional 40

[5] Note that order of contrast is not counterbalanced. Indeed, the phoneme condition serves as a baseline measure of memory performance, as well as a way for participants to learn the task. If order was counterbalanced, French participants who would have to begin with the stress contrast might have difficulties learning the task (whereas the same would not hold for Spanish participants).

distractor pairs, used for the phoneme condition, differed only in one phoneme. Five of these were monosyllabic, 22 were bisyllabic, 10 were trisyllabic, and three were quadrisyllabic. The stimuli were divided over two lists, each composed of 56 experimental words, 56 experimental non-words, 20 distractor words and 20 distractor non-words; a list could not contain both a real word and its corresponding non-word.

Participants were instructed to decide as quickly and accurately as possible if the stimuli presented were words or not by pressing one of two labeled buttons. Each participant was tested with just one list. The order of presentation of the stimuli was randomized for each participant. Preceding the test phase, 30 warm-up trials were presented. None of the warm-up items was included in either experimental list, and none of them was a non-word differing from a real word in the position of stress only. During the warm-up phase, feedback was given after each response with the words 'correct' and 'incorrect'. During the test phase, no feedback was given. A trial started with the presentation of a fixation point, an asterisk in the center of the screen, during 500 ms. The auditory stimulus was presented 500 ms after the disappearance of the fixation point. The next trial began 3500 ms after the end of the auditory stimulus. Instructions were given in Spanish.

In all three tasks, participants were tested in fully equivalent laboratory settings in Paris and Barcelona, using the same software package (EXPE; Pallier, Dupoux, & Jeannin, 1997). The entire experiment lasted about an hour.

3.1.4. Participants

All three groups (simultaneous bilinguals, Spanish and French controls) participated in the Experiment. In the first task, one simultaneous bilingual was rejected, due to equipment failure. Two French controls were rejected, because they made too many reversals. One Spanish control was rejected and replaced for the same reason. In the second task, two simultaneous bilinguals and three French controls were rejected, due to too many reversals. One Spanish control was rejected and replaced for the same reason.

Table 2
Mean results (SEs between parentheses) for the sequence recall tasks and the lexical decision task for Spanish controls, simultaneous French–Spanish bilinguals, and French controls. The numbers in italics are used to compute the composite Z-scores presented in Fig. 1.

	Spanish controls	Simultaneous bilinguals	French controls
Sequence recall #1			
Stress error rate	18.8% (2.7)	38.9% (5.1)	48.7% (2.9)
Phoneme error rate	20.6% (2.5)	29.1% (3.0)	24.8% (2.5)
Difference score	*−1.9 (2.2)*	*9.8** *(4.5)*	*23.9**** *(3.3)*
Sequence recall #2			
Stress error rate	19.6% (2.3)	46.9% (5.4)	72.7% (3.0)
Phoneme error rate	25.0% (4.6)	28.6% (3.7)	22.5% (3.2)
Difference score	*−5.4 (3.7)*	*18.4*** *(4.9)*	*50.2**** *(3.8)*
Lexical decision RTs (ms)			
Stress	1001 (19)	1146 (30)	1436 (36)
Phoneme	999 (26)	1126 (36)	1318 (32)
Mean	*1000 (16)*	*1136 (23)*	*1377 (24)*
Error rates			
Stress	5.1% (0.7)	14.7% (2.6)	41.2% (2.4)
Phoneme	1.8% (0.6)	4.2% (1.3)	10.7% (1.3)
Mean	*3.4% (0.5)*	*9.5% (1.5)*	*25.9% (1.8)*

* $p < .05$.
** $p < .005$.
*** $p < .001$.

3.2. Results

The results for the three groups (simultaneous bilinguals, French, and Spanish) are presented in Table 2. For the two sequence recall tasks, we show the mean error rates in the stress and the phoneme conditions as well as the mean difference scores (defined as stress-minus-phoneme error rate); for the lexical decision task, we show the mean reaction times and the mean error rates in the stress and the phoneme conditions, as well as the overall mean reaction times and error rates in these conditions.

3.2.1. Group analyses

For the three tasks, we used ANOVAs with the factors Group and Condition to compare the stress and phoneme error rates of the three populations: bilinguals, Spanish controls and French controls.

In the first task, the ANOVA yielded main effects of Group ($F(2, 76) = 10.7$, $p < .001$, $\eta^2 = .131$) and Condition ($F(1, 76) = 42.4$, $p < .001$, $\eta^2 = .118$), and a significant interaction between Group and Condition ($F(2, 76) = 13.4$, $p < .001$, $\eta^2 = .075$). This interaction is due to the fact that the Spanish controls made slightly more errors in the phoneme condition (stress: 18.8%, phoneme: 20.6%; $F < 1$, $\eta^2 = .036$), whereas the French controls made significantly more errors in the stress condition than in the phoneme condition (stress: 48.7%, phoneme: 24.8%; ($F(1, 36) = 52.4$, $p < .001$, $\eta^2 = .593$) and the simultaneous bilinguals showed an intermediate pattern (stress: 38.9%, phoneme: 29.1%; $F(1, 21) = 4.8$, $p < .05$, $\eta^2 = .186$). To compare the effect of condition across groups, we computed individual stress-minus-phoneme difference scores, and ran two one-way ANOVAs, one comparing the simultaneous bilinguals to the Spanish controls, one comparing them to the French controls. Both comparisons were significant ($F(1, 40) = 5.1$; $p < .03$, $\eta^2 = .114$, and $F(1, 57) = 6.6$; $p < .02$, $\eta^2 = .104$, respectively).

In the second task, the ANOVA yielded main effects of Group ($F(2, 74) = 17.1$, $p < .001$, $\eta^2 = .142$) and Condition ($F(1, 74) = 125.0$, $p < .001$, $\eta^2 = .239$), and a significant interaction between Group and Condition ($F(2, 74) = 46.3$, $p < .001$, $\eta^2 = .177$). This interaction is due to the fact that the Spanish controls made slightly more errors in the phoneme condition (stress: 19.6%, phoneme: 25%; $F = 2.1$, $p > .1$, $\eta^2 = .098$), whereas the French controls made significantly more errors in the stress than in the phoneme condition (stress: 72.7%, phoneme: 22.5%; $F(1, 35) = 174.6$, $p < .001$, $\eta^2 = .833$) and the simultaneous bilinguals showed an intermediate pattern (stress: 46.9%, phoneme: 28.6%; $F(1, 20) = 14.2$, $p < .002$, $\eta^2 = .415$). As above, individual stress-minus-phoneme difference scores were used to run two one-way ANOVAs, one comparing the simultaneous bilinguals to the Spanish controls, the other comparing them to the French controls. Both comparisons were significant

($F(1, 39) = 14.7$; $p < .001$, $\eta^2 = .274$, and $F(1, 55) = 26.2$; $p < .001$, $\eta^2 = .323$, respectively).

In the third task, we used two ANOVAs, one by subjects and one by items, with average error rates as the dependent variable. We found main effects of Group ($F1(2, 79) = 84.2$, $p < .001$, $\eta^2 = .335$; $F2(2, 380)=574.2$, $p<.001$, $\eta^2 = .396$) and Condition ($F1(1, 79) = 231.3$, $p < .001$, $\eta^2 = .287$; $F2(1, 190) = 190.2$, $p < .001$, $\eta^2 = .183$), and a significant interaction between Group and Condition ($F1(2, 79) = 49.7$, $p < .001$, $\eta^2 = .123$; $F2(2, 79) = 157.0$, $p < .001$, $\eta^2 = .108$). This interaction is due to the fact that the French controls made significantly more errors in the stress condition than in the phoneme condition (stress: 41.4%, phoneme: 10.7%; both $p < .001$, $\eta^2 = .894$), the Spanish controls had an effect in the same direction although of a smaller magnitude (stress: 5.1%, phoneme: 1.7%; $p < .01$, $\eta^2 = .350$), and the simultaneous bilinguals showed an intermediate pattern (stress: 14.7%, phoneme: 4.2%; $p < .002$, $\eta^2 = .393$). To compare the effect of condition across groups, we computed individual stress-minus-phoneme difference scores, and ran one-way ANOVAs, one to compare the simultaneous bilinguals to the Spanish controls, the other comparing them to the French controls. Both comparisons were significant ($F(1, 41) = 4.5$, $p < .04$, $\eta^2 = .099$, and $F(1, 60) = 38.8$, $p < .001$, $\eta^2 = .393$, respectively).

The analysis of the reaction times was not performed because of the very high error rates of the late learners. However, the numerical pattern was very similar, with the reaction times of the simultaneous bilinguals falling in between those of the Spanish and the French controls.

3.2.2. Analyses of the distribution of individual scores

We computed individual composite stress 'deafness' indices across the three tasks as follows. We extracted the individual mean stress-minus-phoneme difference scores from the three tasks. As these scores were expressed on different scales, we normalized them to Z-scores using the Spanish control group as the reference population. The three scores were significantly correlated with one another (Exp. 1 and Exp. 2: $r = .56$, $p < .007$; Exp. 2 and Exp. 3: $r = .55$, $p < .007$; Exp. 1 and Exp. 3: $r = .56$, $p < .008$), confirming that the three tasks measure overlapping processes. In order to derive an estimate of the common process, we defined the individual stress 'deafness' index as the first component in a Principal Component Analysis (PCA) of the three Z-scores. This component accounted for 73% of the variance, and the loadings for the three tasks were, respectively, 0.35, 0.34, and 0.31, showing approximately equal contributions of each of the tasks. Fig. 1 shows the distribution of the composite stress 'deafness' indices across the three participant groups.

In order to determine whether the distribution of the stress 'deafness' index in simultaneous bilinguals is best described as monomodal or as bimodal, we fitted the data using maximum likelihood, which is reliable even for small sample sizes. Specifically, we used two distribution models, one with a single Gaussian, and one with a mixture of two Gaussians (log likelihoods: −46.5 and −39.5, respectively). We then applied Akaike's Information Criterion (Akaike, 1974) in order to compare the fits of these two models while taking into account the difference in degrees of freedom (2 versus 5, respectively).[6] The bimodal distribution was found to fit the data better than the monomodal distribution (Delta AIC = 7.97, $X^2(3) = 13.9$, $p < .003$). Hence, the simultaneous bilinguals' performance appears to be best described by a bimodal distribution. Moreover, this bimodal distribution is very similar to the one obtained by pooling together the data from the Spanish and French controls. Indeed, we found that a bimodal mixture model whose modes correspond to those of the control populations fit the simultaneous bilingual data better than a monomodal distribution (AIC = 9.42; $X^2(1) = 7.4$, $p < .01$), but not better than the maximum likelihood derived above (AIC = −1.45; $X^2(4) = 6.5$, $p > .1$). The maximum likelihood bimodal distribution enabled us to estimate that 43% of the simultaneous bilinguals were in the Spanish-like mode and 57% in the French-like mode (with a standard error of 12% in this estimate). The same computation carried out with the control-based mixture model gave an estimate of 58% of Spanish-like bilinguals and 42% of French-like bilinguals (with a standard error of 12%).[7] The values for means, standard devia-

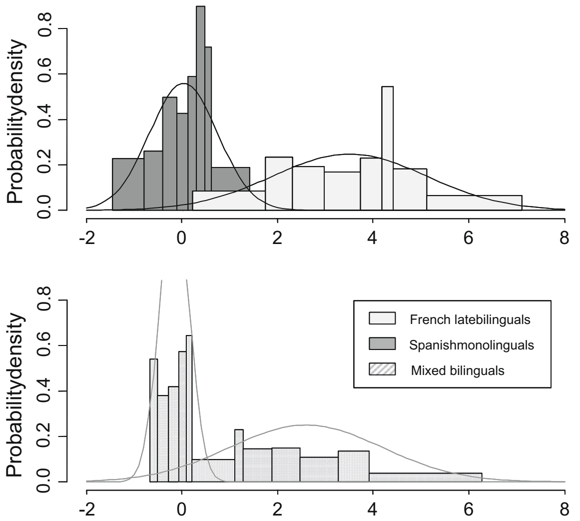

Fig. 1. Probability densities of the distributions of the composite stress 'deafness' indices of Spanish and French controls (top panel) and simultaneous bilinguals (bottom panel).

[6] For model comparisons, AIC is defined as $2k - 2L$, where k is the number of degrees of freedom of the model, and L the log-likelihood. The model with highest AIC is selected as the 'best' one. Here, to compare models two, $M1$ and $M2$ (with $L1$ and $L2$ as likelihoods and $k1$ and $k2$ as degrees of freedom), we report the difference in AIC between them (Delta AIC). We also report the log-likelihood ratio test for that model comparison, i.e. we compute $-2\log(L1/L2)$ which is approximated by a X^2 distribution, with $2(k1 - k2)$ degrees of freedom.

[7] For comparison purposes, we reanalyzed the data of the simultaneous Spanish–Catalan bilinguals tested by Sebastián-Gallés et al. (2005) in the same manner. We found that, like in our study, a bimodal mixture model was more likely than a monomodal model (AIC = 10.2, $X^2(3) = 16.2$, $p < .001$). However, unlike in our study, the maximum likelihood bimodal model differed from the bimodal model computed from the control populations (AIC = 15.5, $X^2(4) = 23.5$, $p<.001$). The reason of this discrepancy is that, as noted by the authors, the 'Catalan' mode of the simultaneous bilinguals is more error prone than that of the Catalan control group ($t(66) = 4.4$, $p < .001$, Cohen's $d = 1.1$), whereas the 'Spanish' mode does not differ from that of the Spanish control group ($t(49) < 1$; $p > .1$).

Table 3
Characteristics of the three models used to fit the stress 'deafness' indices of the simultaneous bilinguals. SEs are shown between parentheses.

	Monomodal model	Maximum likelihood bimodal mixture model	Control-based bimodal mixture model
Loglikelihood	−46.5	−39.5	−42.8
$Mean_1$	1.37 (0.38)	−0.17 (0.10)	0.035 (0.156)
SD_1	1.83 (0.27)	0.30 (0.07)	0.70 (0.11)
$Mean_2$	n.a.	2.52 (0.53)	3.51 (0.26)
SD_2	n.a.	1.63 (0.35)	1.59 (0.18)
p	n.a.	0.43 (0.12)	0.58 (0.12)

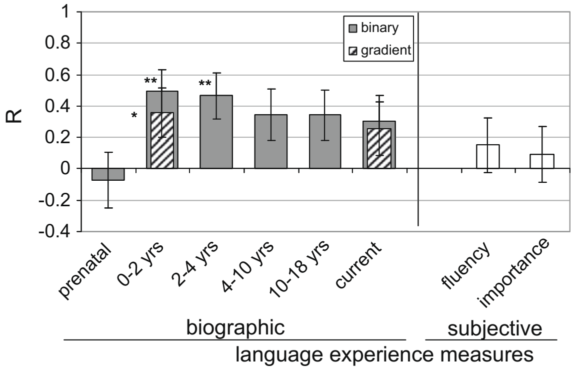

Fig. 2. Correlation coefficients between composite Z-scores and 10 measures of language experience, eight biographic and two subjective. Two biographic measures are gradient (language exposure during the first two years of life, and current language use), the remaining ones are binary (prenatal exposure, country of residence at 0–2 years, 2–4 years, 4–10 years, and 10–18 years, and current country of residence). The error bars represent SEs.

tions and mixture parameters for the three models are provided in Table 3.

3.3.3. Correlation analyses

Finally, we investigated the origin of language dominance in the simultaneous bilinguals by analyzing the correlation of their composite stress 'deafness' indices with 10 independent measures of language experience, eight biographic and two subjective ones. Six of the biographic measures were binary: prenatal exposure (i.e. language of the mother), country of residence at 0–2 years, 2–4 years, 4–10 years, and 10–18 years, and current country of residence (see Table 1); for all of these, Spain was coded as 0 and France as 1. The remaining two biographic measures were gradient. One concerned the language exposure during the first two years of life. We defined this measure as the mean of two highly correlated measures shown in Table 1 (after normalization): early exposure (percentage of French relative to Spanish) as estimated by the parents and as inferred from the participants' interviews ($r = .93$, $p < .0001$). The other gradient biographic measure concerned current language use (inferred from the participants' interviews as a percentage of French use). Finally, the two subjective measures were gradient as well and we collected on a 10 points subjective scale for each language separately, and then converted into a Spanish–French difference score. The first measure was importance, the second one fluency, which we defined as the mean of pronunciation, grammar, and vocabulary (see Table 1). The correlations of the latter three measures are .82 for pronunciation and grammar, .76 for pronunciation and vocabulary, and .83 for grammar and vocabulary (all $p < .0001$).

Fig. 2 shows the correlation coefficients of the 10 measures (one for each language experience measure). Two things are worth noting. First, three measures were significant: country of residence at 0–2 ($r = .49$, $p < .003$) and at 2–4 years ($r = .46$, $p < .005$), and language exposure during the first two years of life ($r = .36$, $p < .04$). After Monte-Carlo resampling ($N = 10^5$) to correct for multiple comparisons (Efron & Tibshirani, 1993), only the first two of these yielded a significant correlation ($p < .025$).[8] Second, two other measures yielded very low correlation coefficients: prenatal exposure ($r = -.07$, $p > .1$) and subjective importance ($r = .09$, $p > .1$). A post hoc test using Monte-Carlo resampling ($N = 10^5$) showed that prenatal exposure had a significantly smaller correlation coefficient than country of residence at 0–2 ($p < .04$) and 2–4 years ($p < .04$). The same comparisons with importance were marginally significant ($p = .09$).

4. General discussion

In this study, we probed simultaneous French–Spanish bilinguals for the perception of Spanish lexical stress using three tasks, two short-term memory encoding tasks, and a speeded lexical decision. With all three tasks, we found that the performance of the group of simultaneous bilinguals was intermediate between that of native speakers of Spanish on the one hand and that of French late learners of Spanish on the other hand. A composite performance measure computed over the results of the three tasks revealed that the overall performance of the simultaneous bilinguals is best fitted by a bimodal distribution that corresponds to a mixture of the performance distributions of the two control groups. The simultaneous bilinguals were about equally divided over these two distributions. Finally, correlation analyses showed that the variables explaining language dominance are linked to early language exposure.

The basic finding that our group of simultaneous bilinguals is not comparable to either a group of native speakers of Spanish or a group of French late learners of Spanish is in agreement with the earlier studies of on-line perception by simultaneous bilinguals (Cutler et al., 1989; Cutler et al., 1992; Sebastián-Gallés et al., 2005; see also Peperkamp

[8] The following procedure was used: chance distributions were constructed by randomly permuting the participants' composite stress 'deafness' indices. For each of these chance distributions, the correlation coefficients for each of the 10 languages dominance measures were computed. The largest correlation coefficient was then compared to the distribution of all largest correlation coefficients obtained on 100,000 iterations, and its rank was converted into a probability. This entire procedure was repeated on nine out of the ten measures, after elimination of the best measure, and so on, until all of the measures were processed. This procedure amounts to performing a Bonferroni correction when all the variables are statistically independent; however, it is much more sensitive than a Bonferroni correction when the variables are correlated, as is the case here.

et al., 1999). However, our study is the first one to demonstrate a bimodal distribution in simultaneous bilinguals. This reinforces the claim by Cutler et al. (1989), Cutler et al. (1992) and Sebastián-Gallés et al. (2005) that simultaneous bilinguals can have only one language that is processed in a native- or near-native-like fashion, at least as far as phonological perception is concerned. Note, however, that we cannot exclude that some or all bilinguals who have native-like performance in Spanish also have (near-)native-like performance in French. In order to fully test the above-mentioned claim it is indeed necessary to probe the performance of simultaneous bilinguals on *two* contrasts, one specific to one of the bilinguals' language, the other one specific to the other language. The prediction is that there is an anti-correlation across the two languages, with no individual performing in the same modes for the two languages. We only tested performance with one contrast, and hence our data do not bear on this issue.[9]

We found that the performance of the Spanish-dominant bilinguals is undistinguishable from that of Spanish monolinguals. This result is in agreement with Cutler et al. (1989, 1992), but contrasts with the finding of Sebastián-Gallés et al. (2005) that among Spanish–Catalan simultaneous bilinguals, the performance of the Catalan-dominant ones is slightly worse than that of native Catalan speakers. We also found that the performance of the French-dominant bilinguals is very similar to that of French late learners of Spanish, who have had only a few years of experience with this language: the former indeed processed stress only marginally better than the latter (and for the late learners, no significant differences were found by Dupoux et al. (2008) among beginners, intermediate and advanced learners). This result is in agreement with both Cutler et al. (1989, 1992) and Sebastián-Gallés et al. (2005), as well as with a hypothesis from the L2 acquisition literature, stating that the procedures for learning a second language are similar across all ages and that there might be no sensitive period for second language learning (Birdsong, 2005; Hakuta et al., 2003). The acquisition of the first language is, however, substantially different from that of a second language, in that L1 acquisition is optimal only within a relatively limited time-window. As far as on-line phonological perception is concerned, our tentative proposal, congruent with all of the available data, is that: 1. the acquisition of the dominant language depends on optimized language learning strategies that can only apply to a single language and that are available for a limited time, and 2. the acquisition of a non-dominant language is dependent on general compensation strategies for which there is no sensitive period, but rather, a slow decline with age.

Our finding that about half of the simultaneous French–Spanish bilinguals have difficulties perceiving Spanish lexical stress is particularly striking in light of the fact that Spanish stress can be encoded without obvious associated costs for the processing of French. That is, whereas both languages use suprasegmental cues to mark prosodic boundaries and grammatical information, only Spanish has lexically contrastive stress; French has been described as either having no stress at all or as having stress on phrase-final syllables (Rossi, 1980). This situation should be compared to those studied by Cutler et al. (1989, 1992) and Sebastián-Gallés et al. (2005), where the two languages of the bilinguals are mutually incompatible with respect to the feature under scrutiny: Cutler et al. (1989, 1992) used a fragment detection task and a word spotting task to probe for the nature of segmentation strategies, that is, strategies used for extracting potential word boundaries from the continuous signal. Crucially, segmentation can be either syllable-based (optimized for French input) or stress-based (optimized for English input), but not both. Likewise, Sebastián-Gallés et al. (2005) used a speeded lexical decision task on Catalan words and non-words differing in the vowel /e/ vs. /ɛ/, Spanish having only one mid-front vowel. Under the assumption that the recognition of speech sounds is based on phonetic prototypes (Kuhl, 1991), the Spanish and Catalan vowel systems are mutually incompatible, because the prototype of Spanish /e/ falls in between those of Catalan /e/ and /ɛ/ (Bosch, Costa, & Sebastián-Gallés, 2000; see also the competition model, Hernandez, Li, & MacWhinney, 2005; MacWhinney, 2005). Hence, as argued by their authors, it might be the case that competition between incompatible systems is what explains why simultaneous bilinguals in these previous studies were found to be impaired in one of their languages.

Finally, our data shed new light on the factors that might lead to language dominance in simultaneous bilinguals. Cutler et al. (1992) tested four variables: country of residence, language of the father, language of the mother[10], and subjective preference. They found no particular effect of the first three variables (although language of the mother was mentioned as coming closest to being significant), but an effect of subjective preference. Sebastián-Gallés et al. (2005) found an effect of the language of the mother; the other variables that they tested, current language exposure and use at home and outside home, did not yield a significant effect. It should be noted that in their study, language of the mother correlated with amount of early exposure. In our group of simultaneous bilinguals, by contrast, language of the mother and early exposure were uncorrelated. We found that language dominance is best predicted by the amount of early exposure (prenatal exposure, $r = .108$, ns). Contrary to Cutler et al., we found no effect of either one of our subjective measures, i.e. fluency and importance.[11] This is in agreement

[9] Cutler et al. (1989), focusing on French–English simultaneous bilinguals, reported an anti-correlation for English-dominant participants but not for French-dominant ones (the latter displayed native-like performance in both languages). However, in a follow-up with the same bilinguals, Cutler et al. (1992) found that on a different task, French-dominant bilinguals do have non-native performance in English, in agreement with the anti-correlation prediction. The study of Sebastián-Gallés et al. (2005), like the present one, tested the performance in one language only.

[10] In contrast to the participants in the present study and the one by Sebastián-Gallés et al. (2005), the bilinguals did not necessarily have one French-speaking and one English-speaking parent; a quarter of them indeed had parents with the same native language.

[11] Cutler et al. measured subjective preference by asking which language participants would want to keep if they had to lose one. Informal interviewing of our participants revealed that they had a difficult time disentangling subjective preference from importance in their current lives. It would probably be better to use more implicit measures of preference instead of an explicit questionnaire that is always susceptible to response biases (Greenwald, McGhee, & Schwartz, 1998).

with a study by Flege, MacKay, and Piske (2002) on early and late bilinguals, that also reported no correlation of language dominance with a subjective, self-rating, measure.

Overall, the data of the three studies are compatible with a primary role of early exposure for on-line phonological perception. This is in accordance with the finding that already at nine months of age monolingual French- and Spanish-learning infants exhibit differences with respect to the perception of stress. Specifically, French- but not Spanish-learning infants fail to perceive stress contrasts (Skoruppa et al., 2009). The three studies are also compatible with the hypothesis that there is no independent role of the language of the mother, suggesting that very early (i.e. prenatal) exposure is irrelevant to the selection of the dominant language. Furthermore, all three studies are compatible with the absence of a large role of current use, in accordance with the proposal that the dominant language cannot be changed after a sensitive period, at least as far as on-line perception is concerned (Parlato-Oliveira, Christophe, Hirose, & Dupoux, in press).

It should be noted that the scope of the present conclusions concerning the origin of language dominance is limited by the fact that some of our biographic variables are correlated, since most bilinguals did not move from one country to another during childhood and adolescence. A larger and especially more diverse population is needed to further refine the impact of variables associated with language dominance. Note also that none of the explored variables has a very strong explanatory power: our best variable (country of residence at 0–2 years) only accounts for 32% of the variance. This may mean that language dominance is determined partly by a number of other factors, including idiosyncratic ones related to personal history, whcih are hard to measure.

To conclude, our results buttress the claim that simultaneous bilinguals are functionally close to monolinguals in their dominant language (Grosjean, 1989), and to late learners in the other language. Future research, using more participants and testing both different contrasts and different processing levels, might shed light onto the nature of language dominance in simultaneous bilinguals as well as the factors that govern the selection of the dominant language.

Acknowledgements

Research for this paper was funded by grants from the Agence Nationale de la Recherche (ANR-05-BLAN-0065-01, ANR NT09_651682 SOCODEV), the European Commission (FP6-STREP "Neurocom" 012738) the Spanish Ministerio de Ciencia e Innovación (SEJ2007-60751), and the CONSOLIDER 2010 Program (CSD2007-012), as well as by the "ICREA Academia" prize for excellence in research awarded by the Generalitat de Catalunya to N.S. We are grateful to Katherine White for help in running the participants in Paris, and to Eduardo Navarrete for testing the participants in Barcelona. We thank Jacques Mehler for comments and discussion.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109–127). New York: Oxford University Press.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 99–116). Amsterdam: John Benjamins.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12*, 189–222.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *The Journal of the Acoustical Society of America, 54*, 421.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilingualism. *Nature, 340*(6230), 229–230.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation in bilinguals. *Cognitive Psychology, 24*, 381–410.

Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36*, 406–421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America, 110*, 1608–1618.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition, 106*, 682–706.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Flege, J., MacKay, I., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23*, 567–598.

Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*, 467–491.

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*, 78–104.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Grosjean (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3–15.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence. A test of the critical-period hypothesis for second-language acquisition. *Psychological Science, 14*, 31–38.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Science, 9*, 220–222.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93–107.

Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science, 97*, 11850–11857.

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). New York: Oxford University Press.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). Expe5: An expandable programming language for on-line psychological experiments. *Behavior Research, Methods, Instruments and Computers, 29*, 322–327.

Parlato-Oliveira, E., Christophe, A., Hirose, Y., & Dupoux, E. (in press). Plasticity of illusory vowel perception in Brazilian–Japanese bilinguals. *Journal of the Acoustical Society of America.*

Peperkamp, S., Dupoux, E., & Sebastián-Gallés, N. (1999). Perception of stress by French, Spanish, and bilingual subjects In *Proceedings of Eurospeech '99* (Vol. 6, pp. 2683–2686).

Rossi, M. (1980). Le français, langue sans accent? [French: A non-stressed language?]. In I. Fonagy & P. Léon (Eds.), *L'accent en français contemporain [stress in modern French]* (pp. 13–51). Paris: Didier.

Sebastián-Gallés, N., Echeverria, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., et al. (2009). Language-specific stress perception of stress by nine-month-old French and Spanish infants. *Developmental Science, 12*, 914–919.

Weber-Fox, C., & Neville, H. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research, 44*, 1338–1353.

Yeni-Komshian, G., Flege, J., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism, Language and Cognition, 3*, 131–149.