The Production and Perception of Voice Onset Time in English-Speaking Children Enrolled in a French Immersion Program

Nicole Netelenbos, Fangfang Li
Department of Psychology, University of Lethbridge, Lethbridge, Canada
nicole.netelenbos@uleth.ca, fangfang.li@uleth.ca

Abstract

The present study investigated the acquisition of French stop consonants in English-speaking children who attend a French immersion elementary school in Alberta, Canada. Both languages have voiced and voiceless stops, but in English the distinction is phonetically realized as short-lag vs. long-lag VOTs, while in French it is as lead vs. short-lag VOTs. Examining the development of stops in this group of children can shed light on the interaction between L1 and L2 when L2 is acquired in an immersion setting. Children in grades 1, 3, and 5 first participated in a word-repetition task, repeating back words beginning with /b/ and /p/. They were then tested on their identification of stop categories over a VOT series ranging from -70 to +70ms, with 10-ms differences among the tokens. Results from the production experiment indicated a clear separation of the two languages in /p/, but not in /b/. The perception results suggest consistent identification in both English stops, the French /p/, but not the French /b/. The results are discussed in terms of the cross-language phonetic differences, the particularity of the learning environment, and the impact of dominant language in society.

Index Terms: VOT, speech perception, speech production, French immersion, second language learning.

1. Introduction

French immersion is a form of bilingual education in which children whose first language is English receive instruction in French in all course subjects, and are thus "immersed" in a French-speaking environment in a school setting. French immersion programs have become increasingly popular since 1970, especially in Western Canada: more than 342,000 students attended immersion programs in elementary and secondary schools in 2011, and the enrollment continues to rise steadily [1]. The immersion form of education creates a new generation of Canadian population who are "bilingual", but in a very different sense from the traditional bilinguals: they do not receive second language (L2) input from their home environment, nor is the L2 they are learning the dominant language in the society. To date, limited research has been dedicated to exploring early L2 acquisition in children via a French immersion program.

The present paper aims to describe the phonetic development of French in immersion children, with the intention of shedding light on the effects of both the nature of L2 input and a minority L2 environment on second language learning. The sounds of focus were labial stops /b/ and /p/. Stop consonants were chosen given that both English and French contrast voiced and voiceless categories, but the phonetic realizations of the two categories differ between the two languages as assessed through voice onset time. Voice onset time (VOT) refers to the time elapse between the release of the stop occlusion and the onset of vocal cord vibration in the following vowel. In English, the voiceless stops in word initial position are aspirated and have long-lag VOT values, while the voiced stops are unaspirated and are realized as short-lag VOTs. In contrast, French voiceless stops are unaspirated and have short-lag VOTs, while voiced stops consist of heavy pre-voicing which typically result in lead (negative) VOTs [2]–[4].

When learning a second language, the speed of acquisition and the proficiency of utilizing the L2 are dependent on a number of factors, including the age of learning [5], the quality of the L2 input [6], length of exposure [7], daily uses of L1 and L2 [8], the status of L1 and L2 in the society (majority vs. minority) [9], and speakers' motivation and attitudes [10]. As French immersion provides a unique case of L2 learning environment, our study specifically addresses three research questions: 1) What is the developmental pattern of the French voicing contrast in children of different grades and do children achieve a more native-like proficiency as the amount of exposure increases? 2) Are children able to maintain two separate language systems at different stages of learning? 3) Is there a link that exists between production and perception during this developmental process of French acquisition?

Many studies have shown that the two language systems of bilingual speakers may interact with each other in speech production [11]–[14] and perception [6], [15]. One of the prominent factors attested in mediating the manifestation of such interaction is age of L2 learning: early L2 exposure generally facilitates the organization of two separate language systems whereas late L2 learning may result in an undifferentiated system [5], [11], [16]–[17]. For instance, Baker and Trofimovich [11] found that early Korean/English bilinguals maintained two separate vowel systems between their L1 and L2 and that this was also evident in the early stages of L2 learning. Conversely, the L2 vowel production of late bilinguals reflected a compromise between L1 and L2.

It is difficult to predict whether French immersion children will maintain two distinct language systems or not, and if not, the nature of interaction. This is because, although these children are exposed to French fairly early in life (starting in kindergarten at age 5) and are immersed in a French environment approximately 7 hours a day and five days a week, the nature of L2 input is not diverse nor is it always authentic (mainly from a single instructor throughout the course of the school year, who may or may not be a native speaker of French). Further, English exists as the dominant language outside of the school setting, with limited opportunities of utilizing the French language in other contexts.

The present study served to explore both the speech production and perception of French stop consonants, not only because both are important aspects of L2 learning, but because previous research has disagreed in regard to whether the two aspects develop separately in second language acquisition or

whether a relationship between the two exists. Flege's [18] speech learning model (SLM) holds that the production of L2 sounds may be inaccurate if the sounds are not correctly perceived by the L2 learners, and several of his studies have provided support for this relationship (e.g. [19]–[20]). Conversely, other studies have found these two processes to be autonomous. For example, Zampini [21] investigated the Spanish (L2) production and perception of stops /p/ and /b/ among adult native English speakers and found no correlation between their VOT production and perceptual abilities.

In the study to follow, VOTs of the voiced and voiceless stops /b/ and /p/ were examined in children enrolled in a French immersion program where both languages (English and French) were tested in a speech production and speech perception experiment. Analyses were performed in children of different grade levels to examine the developmental trend. The VOT values for English and French were also compared to determine whether a possible interaction between the two languages would be revealed. Finally, production and perception results were compared to explore whether a relationship between the two would be observed.

2. Methods

2.1. Participants

A total of 42 French immersion students residing in Alberta, Canada ranging from 6-10 years old in age were included in the study (11 grade 1s of mean age = 6.44, 14 grade 3s of mean age = 8.61, and 17 grade 5s of mean age = 10.45). All participants were native speakers of English learning French through the immersion program and spoke no other languages. All students had learned French prior to the age of 6 and all parents reported their children as normal hearing with no known language, learning or speech delays.

2.2. Experiment 1: Speech Production

The speech production experiment consisted of a word-repetition task where participants were shown images on a computer monitor in a quiet room whilst simultaneously hearing the computer pronounce a word.

The program *Show & Play* [22] was used to display visual stimuli while playing auditory stimuli through a speaker (Logitech Z205, model: S-00094) at the same time. Speech stimuli consisted of pre-recorded speech from a monolingual English speaker for the English component and a native French speaker for the French component. The target labial sounds /b/ and /p/ containing target vowels /i/, /u/, /æ/ (English) or /i/, /u/, /a/ (French) were examined. Children produced a total of 18 words in English (9 with initial stop consonant /b/) and 18 words in French (9 with initial stop consonant /b/). Participants were given verbal instructions prior to the commencement of the task in which they were asked to look at the picture on the computer screen and repeat the word back into the microphone (Shure SM87A) after the word was finished playing. French and English sessions were performed on different days. Children's speech production was recorded using a Marantz flashcard recorder (model: PMD661).

The speech stream was then segmented and events such as bursts and voice onset were marked using Praat [23] to permit extraction of VOT for acoustic analysis. Mispronunciations and repetitions were excluded from analysis (a total of 1.3% for French and .08% for English).

2.3. Experiment 2: Speech Perception

Speech perception was measured by means of a forced choice identification task. During the task, images associated with the target words would appear on the computer screen and then a word synthesized using natural speech would play over the speakers.

The English minimal pair consisted of the words *bear/pear* and the French minimal pair consisted of the words *boule/poule*. Natural speech from a native English speaker and native French speaker were recorded and edited to range on a VOT continuum separated by 10 ms steps ranging from -70 ms to +70 ms VOT yielding a total of 15 tokens per language. In addition, approximately 10% of the stimuli were included as repetitions to gauge intra-subject reliability.

Manual editing for the English stimuli was generated using the original natural sound stimulus of bear (0ms VOT) as the baseline stimulus. All editing was done at the zero crossing of the waveform. The natural speech recording of 'pear' was used as an extraction token. In order to create the first sound stimulus, the release of the oral occlusion /b/ was cut from 'bear' and replaced with the /p/ release from the original 'pear'. Next, 10 ms slices of aspiration were added successively to create the seven remainder positive VOT stimuli. In order to generate the negative VOT values, prevoicing slices in an increment of 10ms were taken from a prevoiced 'bear' and inserted sequentially to the left of the burst of the baseline stimulus.

Manual editing for the French stimuli was generated using the original natural sound stimulus of 'poule' (20ms VOT). In order to create the first sound stimulus, all aspiration (20ms) was cut from the original 'poule' token and then used as the baseline. The remainder of the stimuli were created in the same way as the English stimuli.

Stimuli were played over a speaker connected to a computer in a quiet room using the program *E-prime v 2.0* [24]. All stimuli were randomized over the trials for each participant. Visual stimuli were displayed on the computer monitor and the participant was instructed to use the mouse to click on the corresponding image of the word that they heard (e.g. *boule* or *poule*). Children were shown the images prior to the experiment to ensure that they were familiar with the corresponding words. Visual stimuli were counterbalanced between the first and second half of the experiment. Participants were required to achieve 50% or greater on both the intra-subject reliability measure and extreme value measure, where -70 ms VOT and +70 ms VOT were considered as extreme values.

3. Results

3.1. Age-related developmental change in VOT production

In order to determine whether there exists a difference across the three grades (1, 3, and 5) in the French productions, two repeated measures ANOVAs were conducted, one for /b/ and one for /p/ in French. For both models, the dependent variable was VOT values produced by children and the independent variable was a cross-subject categorical variable Grade (three levels: 1, 3, and 5). No significant differences were revealed in the production of /p/ ($F(2,39) = .38$, $p = .69$) or in the production of /b/ ($F(2,39) = .31$ $p = .74$). The absence of a main effect indicates that the VOT productions in children of the three age groups were not significantly different and thus no developmental trend in regard to VOT was noted.

3.2. L1 and L2 interaction in production

In order to examine whether there would be an interaction between the two language systems in the VOT production of bilingual children, two repeated measures ANOVAs were conducted, one for /b/ and one for /p/. For both models the dependent variable was VOT values produced by children and the independent variables were the cross-subject variable Grade: (three levels: 1, 3, and 5) and within-subject variable: Language (two levels: French vs. English). Findings revealed a significant main effect of language for /p/, ($F(1,39) = 47.89$, $p < 0.01$). Follow-up pairwise comparisons revealed a significant difference between the VOT values of English and French when producing word-initial /p/ within grade 1 ($p < 0.01$), within grade 3 ($p < 0.01$) and within grade 5 ($p < 0.01$). No significant language differences were found when producing word-initial /b/. No significant main effects were revealed for Grade or for Language by Grade. The interaction between L1 and L2 is shown in Figure 1. It is clear from the Figure that children were maintaining separate language systems in their production of the sound /p/, with VOT values for the English /p/ being significantly higher than the French /p/. In contrast, they did not distinguish their English /b/ from French /b/ as the children's realizations of this voiced stop shared similar VOT values around 0 milliseconds (ms).

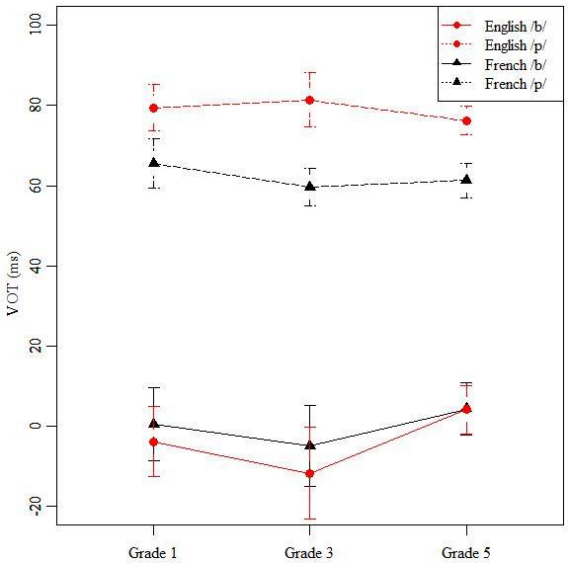

Figure 1. *French immersion children's production of word-initial /b/ and /p/ in both English and French. For each grade, the symbol represents the mean VOT values, and the bars indicate 95% confidence interval.*

3.3. Age-related differences in VOT perception

Figure 2 plots children's identification curves for the /p/-/b/ continuum in both languages. The results indicated that the location of categorical boundary shifted in French when compared to English. Further, the perception of English /p-b/ contrast was far more consistent among the children than that of their French sound perception. The boundary location and slope of identification was very similar across all three age groups in both languages with no developmental trend being observed. In terms of the English (*bear/pear*) stop discrimination, the VOT boundary identification falls around the 25 ms range at the 50% crossover with a fairly steep slope, with the exception of grade 1s whose slope is a bit short of grade 3 and grade 5 (see the left panel in Figure 2). When perceiving the French (*boule/poule*) pair, the VOT boundary is located around the 5 ms range at the 50% crossover revealing a fairly steep slope, although slightly less steep than the children's English perception. The boundary locations exhibited in children's perception are appropriate for each language and are consistent with the perception location reported in monolingual English and monolingual French listeners (e.g. [25]). One thing to note is the imperfect identification of French /b/, suggesting a certain degree of confusion between the prevoiced /b/ with the short-lag /p/ sound in perceiving French. Moreover, the children performed with high accuracy in their identification of the French voiceless token /p/ (see the right panel in Figure 2).

A chi-square test was conducted to evaluate whether there was a significant difference between grades (1, 3, 5) in response frequency for /p/ identification and /b/ identification in the students' French perception. A non-significant chi-square was obtained, χ^2 (2, N = 649) = .094, p = .954. A McNemar test was used to determine whether there would be a significant difference in the number of responses for /p/ identification vs. /b/ identification as a function of language (English, French). Results revealed that response type differed significantly by language with more /b/ identification responses in English than in French, $p < .001$ These results might be interpreted by the difference revealed in boundary location between the two languages where the location of categorical boundary shifted in French when compared to English, eliciting more /b/ identification responses in the English perception task. These findings might also further reflect the difficulty in correctly identifying the /b/ token in the French perception task.

4. Discussion

The objectives set forth in the present study were to describe the French-learning experience of immersion children whose first language is English. By evaluating the VOT values in their labial stop production as well as the delineation of the VOT space in stop perception, it was possible to address 1) whether there exists a developmental trend as children advance through grades 2) whether children are maintaining two separate sound systems between their L1 and L2 and 3) whether production mirrors perception in development.

The findings indicated that student's VOT production patterns and identification functions were not significantly different across age groups. The lack of difference across grades suggests the minimal impact of the students' age and the length of L2 exposure in the immersion program on VOT performance. Across the three grades, in terms of perceptual ability, children attained native-like categorical boundary in French, but have difficulty in identifying the /b/ sound to some extent. In terms of production proficiency in French, children articulated the /p/ sound with significantly smaller VOT values than their English /p/, but falling short of native like French VOT patterns (e.g. [12]). In terms of the voiced stop, overall, their French /b/ is not pre-voiced, as evidenced by the VOT values centering around 0 ms.

The result of a non-native-like French production pattern near the end of elementary schooling (in Grade 5) is consistent with another study of French immersion schooling [26], which investigated comprehension, oral, and written skills of French immersion children, and found that the bilingual children were rated as falling short of native-like French pronunciation skills. We speculate that one important factor contributing to the lack

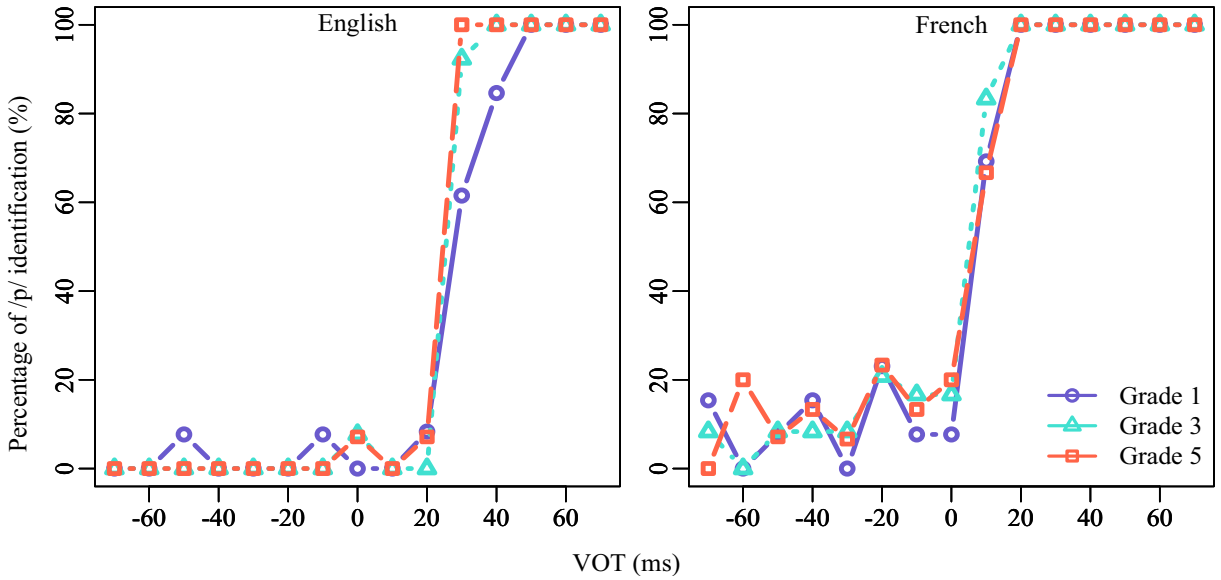

Figure 2. *French immersion children's mean identification scores for the perception of /p/ in both English and French as a function of VOT.*

of native-like VOTs in French immersion children is the minority status of French in Alberta. The only exposure to L2 that the children receive is confined to the classroom setting. Furthermore, in terms of the nature of French input, children are exposed to a single instructor for the entire schooling year, who may or may not be a native speaker of French. Similar to the French immersion children, some teachers also took the French immersion route themselves. In addition, from the author's experience in the immersion classroom and as noted by Swain [27] children will often fall back on their native language when interacting with their peers during group activities. Even though the immersion program strongly encourages students to speak French while in the classroom, many students still return to their L1 to express and explain themselves in a more comfortable way. Evidently, a lack of L2 practice will in turn influence production abilities in L2 acquisition.

With respect to the question of whether children are maintaining two separate phonetic systems, results indicated that a separation exists in the /p/ but not the /b/ between English and French. Even after only a year of exposure, the grade 1s were differentiating their English /p/ from their French /p/. Anderson [28] also noted that children enrolled in an immersion program who were exposed to an L2 for only a short period of time appeared to maintain separate phonological systems in their production of consonants. Our research, however, demonstrated that such a separation between L1 and L2 could occur partially, affecting some sounds only.

Moreover, our results from the perception experiment are harmonious with those of the production results in that, children of all three age groups had no difficulty in correctly identifying the voiceless stop /p/ in both languages, but were less consistent and accurate in identifying the French /b/. The parallel in the production and perception patterns suggests that these two aspects of phonetic development are related. Specifically, the perceptual confusion children are encountering with the French voiced stop /b/ may very well likely be impeding the L2 learners from articulating the sound accurately, which in turn, reinforces the perceptual confusion. The inability to maintain two separate language sound systems in the production of /b/ may also have something to do with the fact that voiced stops can be phonetically realized as short-lag or lead VOTs in English. In other words, the French voiced category is not entirely new phonetically, as it is a form of free variation with short-lag VOT variant, both of which belong to the voiced stop category in English. Therefore, in order to acquire the French voiced category, children need to divide the original VOT range for English voiced stops into two stop categories in French and without sufficient input from the environment, such a division is difficult to establish.

5. Conclusion

Taken together, the findings of the present study suggest that a year experience in French immersion is effective in shifting the perceptual boundary to authentic French, but insufficient for complete native-like production patterns to occur. Also early exposure (at age 5 when children were first immersed in kindergarten) in this case does not guarantee a separate language system for L2. This is in sharp contrast to the research findings in the majority of L2 literature where early acquisition of L2 almost always results in a separate system for L2 that resembles native speakers. Those early learners however receive authentic input from home environment or the L2 is the dominant language in the society. The current study thus demonstrates the importance of the larger social-cultural context as well as the nature of L2 input in the process of second language acquisition.

6. Acknowledgements

We would like to express our sincere gratitude and appreciation to all of the staff of École Agnes Davidson Elementary School in Lethbridge, Alberta, Canada, for all of their help and support in the data collection process. Further acknowledgement extends to the Lethbridge School District No. 51, children who participated in the study, and parents who supported our research.

7. References

[1] Friesen, J. (2013, Jan. 28). *French immersion enrolment skyrockets as a new linguistic category emerges* [Online]. Available: http://www.theglobeandmail.com/news/national/education/French-immersion-enrolment-skyrockets-as-a-new-linguistic-category-emerges/article7935100/?page=all

[2] Fagyal, Z., Kibbee, D. and Jenkins, F., *French: a linguistic introduction.* Cambridge, UK: Cambridge University Press, 2006.

[3] Coveney, A., *The sounds of contemporary French.* Exeter, UK: Elm Bank Publications, 2001.

[4] MacLeod, A. N. A. and Stoel-Gammon, C., "The use of voice onset time by early bilinguals to distinguish homorganic stops in Canadian English and Canadian French", Applied Psycholinguistics, 30:53-77, 2008.

[5] Guion, S., "The vowel systems of Quichua–Spanish bilinguals: An investigation into age of acquisition effects on the mutual influence of the first and second languages", Phonetica, 60:98–128, 2003.

[6] Mackay, I. R. A., Flege, J. E., Piske, T. and Schirru, C., "Category restructuring during second-language acquisition", Journal of the Acoustical Society of America, 110:516-528, 2001.

[7] Bohn, O. S. and Flege, J. E., "The production of new and similar vowels by adult German learners of English", Studies in Second Language Acquisition, 14:131-158, 1992.

[8] Flege, J. E., Bohn, O. S. and Jang, S., "Effects of experience on non-native speakers' production and perception of English vowels", Journal of Phonetics, 25:437-470, 1997.

[9] Mougeon, R. and Beniak, E., "Minority language schooling without home language maintenance: impact on language proficiency", in A. Holmen [Ed], Bilingualism and the Individual, 253-264, Clevedon, 1991.

[10] Oxford, R. and Shearin, J, "Language learning motivation: Expanding the theoretical framework", The Modern Languages Journal, 78:12-28,1994.

[11] Baker, W. and Trofimovich, P., "Interaction of native- and second-language vowel system(s) in early and late bilinguals", Language and Speech, 48(1):1-27, 2005.

[12] Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A. and Hallé, P., "Cross language phonetic influences on the speech of French–English bilinguals", Journal of Phonetics, 36(4):649-663, 2008.

[13] Flege, J. E., "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language", Journal of the Acoustical Society of America, 89(1):395-411, 1991.

[14] Harada, T., "The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program", IRAL: International Review of Applied Linguistics in Language Teaching, 45(4):353-378, 2007.

[15] Guion, S., Flege, J. E., Yamada, R. A. and Pruitt, J.C., "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants", Journal of the Acoustical Society of America, 107:2711–2724, 2000.

[16] Flege, J. E., Schirru, C. and MacKay, I. R. A., "Interaction between the native and second language phonetic subsystems", Journal of Speech Communication, 40:467–491, 2003.

[17] Kang, K. and Guion, S, G., "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals", Journal of the Acoustical Society of America, 119(3):1672-1683, 2006.

[18] Flege, J. E., "Second language speech learning: theory, findings and problems," In W. Strange [Eds.], Speech perception and linguistic experience: issues in cross-language research, 233-277, Timonium, MD: York Press. 1995.

[19] Schmidt, A. and Flege, J. E., "Effects of speaking rate changes on native and non-native production" Phonetica, 52(1):41-54, 1995.

[20] Flege, J. E., MacKay, I. R. A. and Meador, D., "Native Italian speakers' perception and production of English vowels", Journal of Acoustical Society of America, 106(5): 2973-2987, 1999.

[21] Zampini, M. L., "The relationship between the production and perception of L2 Spanish stops", Texas papers in foreign language education, 3(3):85-100, 1998.

[22] Edwards, J. and Beckman, M. E., "Methodological questions in studying phonological acquisition", Clinical Linguistics and Phonetics, 22(12):939-958, 2008.

[23] Boersma, P. and Weenink, D., Praat: Doing Phonetics by Computer, ver 5.3.39, Amsterdam, The Netherlands.

[24] Schneider, W., Eschman, A. and Zuccolotto, A., E-Prime 2.0 software: Psychology Software Tools, Pittsburgh, PA

[25] Caramazza, A., Yeni-Komshian, G. H., Zurif, B. E. and Carbone, E, "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals", The Journal of the Acoustical Society of America, 54:421-428, 1973.

[26] Genesee, F., "A longitudinal evaluation of an early immersion school program", Canadian Journal of Education, 3:31-50, 1978.

[27] Swain, M., "French Immersion Research in Canada: Recent Contributions to SLA and Applied Linguistics", Annual Review of Applied Linguistics, Cambridge University Press, 20:192-212, 2000.

[28] Anderson, R. T., "Phonological acquisition in preschoolers learning a second language via immersion: a longitudinal study", Clinical Linguistics and Phonetics, 18(3):183-210, 2004.