

Lexicality, frequency and stress assignment effects in bilingual children reading Italian as a second language*

STÉPHANIE BELLOCCHI

Aix Marseille University, Aix en Provence

PAOLA BONIFACCI

University of Bologna, Bologna

CRISTINA BURANI

Institute of Cognitive Sciences and Technologies, Rome &

Department of Life Sciences, University of Trieste, Trieste

(Received: May 27, 2013; final revision received: May 28, 2014; accepted: June 02, 2014; first published online 15 December 2014)

In Italian, developing readers exhibit lexicality and frequency effects, and are sensitive to the distributional properties of the language. But how do bilingual children with different ages of first L2 (Italian) exposure and L2 vocabulary sizes read L2 words and pseudowords? Two reading aloud experiments investigated lexicality, frequency and stress assignment effects in fourth- and fifth-grade bilinguals and monolinguals. Naming latencies and pronunciation accuracy were analyzed. In Experiment 1, effects of lexicality and frequency and between-group differences emerged. In Experiment 2, the word frequency effect was confirmed. Late bilinguals, characterized by a smaller L2 vocabulary size, were less accurate than early bilinguals and monolinguals in assigning non-dominant stress. As with monolinguals, lexical information seems to be employed when reading Italian as a second language. Furthermore, bilingual readers are sensitive to the distributional properties of the language. Stress assignment is affected by the L2 lexicon size of second-language learners.

Keywords: lexicality, reading, stress assignment, age of L2 exposure, bilingualism, second-language learners

Introduction

An increasing number of children are being exposed to a reading system in a language that is different from their first language (L1) as a result of migration processes or increased learning opportunities. These children can be defined either as BILINGUAL CHILDREN or as SECOND-LANGUAGE LEARNERS, depending on the type of linguistic exposure they have received (Paradis, Genesee & Crago, 2011). Immigrant children have often been described as performing less well than monolingual children on reading tasks (Slavin & Cheung, 2003). By contrast, parallel research on the so-called BILINGUAL ADVANTAGE has reported enhanced nonverbal skills (e.g. Bialystok, 2009; Bialystok & Feng, 2009; Bonifacci, Giombini, Bellocchi & Contento, 2011; Costa, Hernández & Sebastián-Gallés, 2008). An in-depth analysis of the

cognitive and linguistic processes that intervene in reading in a second language (L2) is imperative, if we are to understand typical and atypical development of this ability. Many studies have shown that learning to read in an L2 is similar, in many aspects, to learning to read in an L1 (for comprehensive reviews, see August & Shanahan, 2006; Genesee & Jared, 2008). Some predictors of reading efficiency, such as phonemic awareness, letter-sound knowledge, working memory, vocabulary, and rapid automatized naming (RAN), have been found to be meaningful indices both for monolingual (e.g. Kirby, Desrochers, Roth & Sandy, 2008) and bilingual reading acquisition (see Genesee, Lindholm-Leary, Saunders & Christian, 2006).

Although most studies in the literature point to similar patterns of reading acquisition in L1 and L2, there are also important differences. These differences may be linked to different sociocultural backgrounds (e.g. teaching methods and materials that are not familiar to children speaking an L2-minority language) (Paradis et al., 2011). Moreover, learning to read in an L2 can be modulated by other specific factors, related to either the reader's prior linguistic exposure or the specific properties of the language that is to be acquired. In the present study, we focused on the interaction between two main factors: age of first bilingual exposure and impact of lexical knowledge. In particular, we sought to test lexicality, frequency, and stress assignment effects in bilingual children reading Italian as an L2. Although a number of

* This work is dedicated to Silvana Contento, formerly Professor at the Faculty of Psychology, University of Bologna. The authors are very grateful to Silvana for sharing her interest in bilingualism and second-language learning with them, and for starting this scientific collaboration.

The authors also thank Ilaria Ceccarelli for her help with the data collection, and Silvia Primativo, Pasquale Rinaldi, and three anonymous referees for their useful comments. Finally, we are very grateful to Bruno Dauvier for his precious assistance with the data analysis.

Since this paper was accepted for publication, the first author's affiliation has changed from Aix Marseille University, Aix en Provence, to Paul Valéry University, Montpellier.

Address for correspondence:

Stéphanie Bellocchi, Paul Valéry University, Route de Mende, 34199, Montpellier, France
stephanie.bellocchi@univ-montp3.fr

studies have addressed these questions in monolingual or bilingual adults, to the best of our knowledge this is the first study to have looked at bilingual children whose L2 is Italian.

Age of first L2 exposure

Based on the definition provided by Kovelman, Baker and Petitto (2008), the term AGE OF FIRST BILINGUAL EXPOSURE (AoE-L2) refers to the age when a bilingual child first begins to receive intensive, systematic, and maintained exposure to his/her new language. In the literature, it is often used interchangeably with the expression AGE OF ACQUISITION, which has been commonly employed to denote the age at which a monolingual individual first starts learning a new or second language. An increasing number of studies have shown that children who are raised in a bilingual environment from birth reach linguistic milestones (e.g. first two-word combinations) in both languages at very similar times to monolingual children (Genesee, 1989; Holowka, Brosseau-Lapr e & Petitto, 2002; Pearson, 1998; Pearson, Fernandez & Oller, 1993; Petitto, Katerelos, Levy, Gauna, Tetreault & Ferraro, 2001). Nevertheless, the debate on the SENSITIVE PERIODS for developing fully proficient linguistic skills in two or more languages is ongoing and controversial, and there are still some misconceptions concerning the relationship between age and L2 learning, as evidenced by Marinova-Todd, Marshall and Snow (2000). Penfield and Roberts (1959) first put forward their CRITICAL PERIOD HYPOTHESIS, subsequently developed by Lenneberg (1967), who defined critical periods for the acquisition of segmental phonology, morphology, and syntax. Then Locke (1997) proposed his OPTIMUM BIOLOGICAL MOMENT VIEW and, in recent years, behavioral data have been enriched by neuroimaging studies. These studies suggest that late bilinguals (LBs), although highly proficient, do not exhibit a native-like pattern of activity in response to a new language (Isel, Baumg rtner, Thr an, Meisel & B chel 2010; Perani, Abutaleb, Paulesu, Brambati, Scifo & Cappa, 2003) whereas early bilinguals (EBs), usually defined as having an AoE-L2 below three years, do not exhibit any significant differences in brain activation patterns in linguistic tasks with respect to monolingual children. However, when we take behavioral indices of L2 proficiency into consideration, the existence and timing of a sensitive period significantly affecting the acquisition of L2 structures become far less clear cut. For example, while Johnson and Newport (1989) highlighted a decline in grammatical judgment in people exposed to English as an L2 after puberty (15 years), when Bialystok and Hakuta (1999) recalculated age–performance correlations, they found that the decline set in after age 20.

While several studies have examined the sensitive period in a variety of domains of linguistic competence, very few have addressed the role of AoE-L2 in reading proficiency. Kovelman et al. (2008) reported that EBs (AoE-L2 within the first three years of life) displayed better L2 reading skills, in terms of accuracy and speed, than LBs (or second-language learners; AoE-L2 four to six years of age), and the reading skills of EBs did not differ from those of monolingual children. Moreover, EBs performed just as well as native speakers of their L1 (Spanish) and L2 (English) on a standardized language competence/expressive proficiency assessment, whereas LBs performed less well in their new language. In the present study, we adopted the definition of EBs used by Kovelman et al. (2008) and further compared EBs and LBs in reading tasks. As previously indicated, we also assessed the impact of AoE-L2 on children’s sensitivity to specific language properties.

Lexical knowledge in reading Italian

Contrary to the claim that word reading in transparent orthographies is accomplished mainly through the nonlexical route (i.e. relying on grapheme–phoneme correspondences), recent studies have shown that the acquisition of reading in transparent orthographies such as Italian is influenced by lexical variables from the earliest stages of development (e.g. Burani, Marcolini & Stella, 2002), in a similar way to what happens in the case of languages with opaque orthographies such as English. Several studies have highlighted lexicality effects (the advantage of reading words rather than nonwords), frequency effects (the advantage of reading high-frequency versus low-frequency words), and stress-assignment effects when reading in Italian (e.g. Bellocchi & Bastien-Toniazzo, 2011; Burani & Arduino, 2004; Burani, Paizi & Sulpizio, 2014; Colombo, 1992; Colombo, Pasini & Balota, 2006; Pagliuca, Arduino, Barca & Burani, 2008; Paizi, De Luca, Zoccolotti & Burani, 2013; Paizi, Zoccolotti & Burani, 2011).

The study of stress assignment in monolingual readers has received increasing attention in recent years, as the attribution of stress is an important process that can tell us how linguistic knowledge is activated in reading tasks. As we explain below, in Italian stress position can vary, and the process of stress assignment in reading requires lexical knowledge, specifically knowledge of the suprasegmental properties of the Italian language. For example, in the word *farina* (“flour”), the stress is placed on the penultimate syllable, whereas in the word *favola* (“tale”), the stress is placed on the antepenultimate syllable.

Most studies examining the relationship between the development of phonological representations and literacy acquisition have focused on the sublexical

segmental level. However, spoken words also carry suprasegmental or prosodic information, defined by fundamental frequency, intensity, duration, and amplitude modulations. At the lexical level, these parameters co-determine the prosodic or stress pattern of words in stress-based languages (e.g. English, Dutch, Spanish) that display a rhythmic alternation between stressed and unstressed syllables (Goetry, Wade-Woolley, Kolinsky & Mousty, 2006). Italian is a transparent language with regular grapheme–phoneme correspondences, but less predictable suprasegmental properties. In words of three or more syllables, the stress can be assigned to either the penultimate or the antepenultimate syllable, but there are no formal rules governing the two cases,¹ and no mark is used in the written form to signal where the stress occurs.² To correctly assign stress, the reader therefore needs to access the lexicon and retrieve the correct pronunciation. Alternatively, the reader can estimate which stress pattern is more likely to be correct based on a probabilistic assessment of the stress properties of his/her language (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Paizi et al., 2011; Sulpizio & Colombo, 2013). For example, in Italian there is a predominance of words with the stress on the penultimate syllable (Thornton, Iacobini & Burani, 1997), which have therefore been defined as *REGULAR WORDS*, or more correctly as words with a *DOMINANT STRESS*, in contrast to words with the stress on the antepenultimate syllable, which have been defined as *IRREGULAR WORDS*, or words with a *NON-DOMINANT STRESS*.

However, stress assignment may interact with another fundamental word property, namely frequency. A number of studies have investigated this interaction with contrasting results. Colombo (1992), in accordance with results obtained for English words with regular versus irregular spelling–sound correspondences (e.g. Andrews, 1992; Seidenberg, Waters, Barnes & Tanenhaus, 1984), found an interaction between stress dominance and word frequency in reading Italian polysyllables: words with a dominant stress were read aloud faster than words with a non-dominant stress, but only when these were of low frequency; that is, they were more likely to be computed through sublexical correspondences. Burani and Arduino (2004) challenged the effect of stress dominance on low-frequency word reading (Colombo, 1992; Rastle & Coltheart, 2000) by showing that stress assignment to low-frequency words can be determined by

the number of words that share the same stress pattern and final orthographic/phonological sequence (referred to as *STRESS NEIGHBORHOOD*) (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; for similar effects in English, see also Arciuli & Cupples, 2006; Kelly, Morris & Verekkia, 1998).

Paizi et al. (2011), together with Sulpizio, Boureux, Burani, Deguchi, and Colombo (2012) and Sulpizio and Colombo (2013), have suggested that beginning and less-skilled readers may be more influenced by stress dominance in reading, possibly because they do not yet possess a stable link between the orthographic representation of a word and its stress pattern, unlike older children or adult readers. The absence of lexical information on the stress position for a given word, with consequent reliance on sublexical processing (Ziegler & Goswami, 2005), may favor the assignment of stress to the penultimate syllable, owing to sensitivity to the statistical distribution of stress patterns, leading children to assign the dominant (i.e. most frequent) stress pattern.

In contrast to the growing literature on stress assignment in monolingual readers, stress assignment in bilinguals has been rarely investigated. Goetry et al. (2006) compared the stress-processing abilities of French and Dutch monolinguals with French–Dutch bilinguals and Dutch–French bilinguals. As well as replicating the *STRESS DEAFNESS*³ effect in native French listeners (Dupoux, Pallier, Sebastian & Mehler, 1997; Dupoux, Peperkamp & Sebastian-Galles, 2001), they observed that the performances of the French–Dutch bilinguals in terms of the accuracy of stress assignment were midway between those of the two monolingual groups. The authors interpreted these data as suggesting that stress-processing abilities can be acquired through L2 exposure in children between four and five years of age. They further observed that stress-processing abilities were significantly correlated with lexical development and reading skills in the French–Dutch bilingual group.

The present study

Taking all these considerations into account, the present study was designed to assess whether bilingual children exposed to Italian as a second language exhibit sensitivity to lexical properties and stress dominance when reading Italian, and whether this sensitivity differs between bilinguals with differing AoE-L2. Thus, the overall purpose of this study was to test lexicality, frequency, and stress-assignment effects in EBs and LBs compared with monolinguals. AoE-L2 is associated with differences in

¹ There is only one rule in Italian for assigning stress to polysyllabic words. This rule refers to the weight of the penultimate syllable: if it is heavy – that is, if it ends with a consonant (e.g. *bi.SON.te* “bison”) – then it has to bear the stress (Krämer, 2009). However, there are exceptions to the rule (e.g. *MAN.dor.la* “almond” or *LE.pan.to* “Lepanto”).

² Fewer than 2% of polysyllabic words bear the stress on the final syllable, and in this case stress is graphically marked (e.g. *libertà* “freedom”).

³ According to the studies of Dupoux et al. (1997, 2001), native speakers of French, as opposed to native speakers of Spanish, exhibit a robust stress deafness, that is, they have considerable difficulty perceiving stress contrasts even if they have learned Spanish as a second language.

L2 lexical knowledge and vocabulary size. Furthermore, the age at which a bilingual child is introduced to a new (or additional) language is thought to impact ultimate dual-language competence and proficiency, such that persons with early exposure to two languages (EBs) achieve greater language mastery than persons with late bilingual exposure (LBs; see, for example, Flege, MacKay & Meador, 1999; Flege, Munro & MacKay, 1995; Johnson & Newport, 1989; Petitto & Kovelman, 2003). We thus predicted that L2 lexicon size, which is assumed to be smaller in LBs, would have a negative impact on their sensitivity to lexical effects and stress assignment. To test this prediction, we conducted two reading aloud experiments: Experiment 1 assessed lexicality and frequency effects, and Experiment 2 the effects of stress assignment and its interaction with word frequency on reading speed and accuracy.

Method

Participants

A total of 66 children aged from eight to ten years (mean age in months: 108.82, *SD*: 6.89) took part in this study. They were all selected and recruited from mainstream primary schools where the Laboratory for the Assessment of Learning Disabilities (LADA) of Bologna University's Department of Psychology was screening for learning disabilities. All the children were in fourth or fifth grade.

The children were divided into three groups:

- (i) Two groups of bilingual children (30), selected on the basis of their age of exposure (AoE) to Italian as a second language (AoE-L2 Kovelman et al., 2008), established by means of a questionnaire completed by parents and teachers (Bellocchi, 2010; see also Contento, Bellocchi & Bonifacci, 2013): 15 EBs with an AoE-L2 before three years and eleven months old; 15 LBs with an AoE above four years. The EBs had all attended kindergarten in Italy while, in order to participate, the LBs had to have been schooled in Italian at least for 3 years [46.7% of them started learning Italian as an L2 at the start of primary school (i.e. first grade), and the remaining 53.3% started learning Italian during the last years of kindergarten]. None of the bilingual children had previously been schooled from first grade in their L1. The selection criteria were:
 - (a) exposure to an L1 different from Italian (L2) within the family context.
 - (b) performance IQ within the normal range (> 85).
 - (c) normal reading abilities defined by scores no more than one standard deviation below the mean on a standardized word reading test.

All the bilinguals were schooled exclusively in Italian. There were 15 girls and 15 boys representing eight native language groups: Albanian (13.3%), Tagalog (33.3%), Moroccan (16.6%), Romanian (20%), Bengali (6.6%), Spanish, Russian, and Mandarin (3.3% each).

- (ii) One group of Italian monolinguals (36; 22 girls and 14 boys) matched with the bilinguals for chronological age and nonverbal IQ. Like the bilinguals, they were all typically developing readers, as assessed by standardized measures.

Screening tests

During the screening phase, each participant's reading, lexical, and cognitive abilities were evaluated using standardized tests. Parental permission was obtained before administering these tests and the subsequent experimental tasks.

Italian reading achievement was tested by means of Tasks 2 (speed in syllables/second) and 3 (accuracy in number of errors in word and pseudoword reading) of the *Batteria per la Valutazione della Dislessia e Disortografia Evolutiva-2* ("Developmental dyslexia and dysorthographia assessment battery"; Sartori, Job & Tressoldi, 2007), which is one of the most widely used Italian tests for assessing word and nonword reading efficiency. Reported test-retest reliability is .77 for reading speed and .56 for accuracy. The third task, *Prove di Lettura MT per la Scuola Elementare* ("MT Reading tests for primary schools"; Cornoldi & Colpo, 1998) consisted in reading aloud a meaningful passage in which speed (syllables/seconds) and accuracy (number of errors) were rated. Passage difficulty varied with grade level, meaning that fourth graders read a different passage from fifth graders. Correlation indices on parallel forms for each grade level ranged from .75 to .87 for accuracy scores and from .94 to .97 for reading speed.

The cognitive assessment took the form of the Kaufman Brief Intelligence Test-2 (K-BIT-2; Kaufman & Kaufman, 2004) an Italian version of verbal tasks adapted in our laboratory for experimental purposes.⁴ This test is composed of vocabulary (verbal knowledge and riddles) and matrix subtests, and provides standardized measures of verbal, performance, and composite full scale IQ. Reliability for native English-speakers (age 4–90 years) is .91 for verbal IQ, .88 for nonverbal IQ, and .93 for composite IQ.

⁴ The Verbal Knowledge subtest measures verbal intelligence and assesses receptive vocabulary and knowledge of general information. The Matrices subtest assesses the ability to solve new problems, perceive relationships and complete visual analogies, and yields a total nonverbal IQ score. Finally, the Riddles subtest measures verbal comprehension, reasoning, and vocabulary knowledge.

Table 1. Monolinguals' and bilinguals' mean scores (standard deviations) on the Italian reading aloud tasks, verbal fluency task (*z* scores) and IQ test (standard scores).

	Monolinguals mean (<i>SD</i>)	Early bilinguals mean (<i>SD</i>)	Late bilinguals mean (<i>SD</i>)	MANOVA
Text reading_speed (<i>z</i> score)	0.4 (0.9)	0.3 (0.6)	-0.5 (0.7)	$p < .05$
Text reading_errors (<i>z</i> score)	-0.4 (1.1)	-0.8 (0.9)	-1.2 (1.7)	<i>ns</i>
Word reading_speed (<i>z</i> score)	0.8 (0.9)	1.2 (0.9)	0.3 (1.2)	<i>ns</i>
Word reading_errors (<i>z</i> score)	0.5 (0.6)	0.4 (0.7)	0.0 (0.7)	<i>ns</i>
Pseudoword reading_speed (<i>z</i> score)	0.5 (0.9)	0.9 (1.1)	0.6 (0.8)	<i>ns</i>
Pseudoword errors (<i>z</i> score)	0.4 (0.8)	0.3 (0.7)	0.1 (0.9)	<i>ns</i>
Verbal fluency (<i>z</i> score)	0.2 (0.9)	-0.4 (0.7)	-0.5 (0.8)	$p < .05$
Nonverbal IQ	92.9 (14.2)	87.4 (9.9)	93.3 (14.1)	<i>ns</i>
Verbal IQ	116.9 (11.6)	105.9 (9.7)	94.6 (16.4)	$p < .05$

Participants' lexical abilities were also assessed by means of a verbal fluency task drawn from the *Batteria per la Valutazione Neuropsicologica 5-11* ("Neuropsychological assessment test for 5-11-year-olds"; Bisiacchi, Cendron, Gugliotta, Tressoldi & Vio, 2005). Participants are asked to verbally produce a set of Italian words that belong to several semantically defined domains (towns, colors, etc.), with a one-minute interval between each domain (defined using a stopwatch). Mean test-retest reliability is .84.

Most of these measures had been standardized on a population of monolingual Italian speakers, which may constitute a limitation when comparing bilinguals. However, they had the advantage of allowing us to analyze these descriptive statistics from a comparative perspective (Uchikoshi & Marinova-Todd, 2012) and to select participants (both monolingual and bilingual) with typical reading and cognitive development.

The entire reading and cognitive assessment with standardized tests lasted approximately 30 minutes.

Sample characteristics: Results

In order to investigate and describe the sample's characteristics, we ran a multivariate analysis of variance (MANOVA) on cognitive and reading abilities, with GROUP (3 levels: monolinguals, EBs, and LBs) as a between-participants factor. Table 1 provides the means and standard deviations of the error rate and speed (*z* scores) in the reading aloud tasks for the monolinguals, EBs, and LBs, together with the verbal fluency *z* scores and IQ scores. Statistics are also reported. Statistical analyses were conducted using the SPSS® program, version 20.0.

Results showed that LBs were significantly slower than EBs ($p < .05$) and monolinguals ($p < .05$) (Bonferroni post hoc test) on text reading ($F(2,66) = 7.38, p < .05, MSE = 0.61$). No significant differences were found for

accuracy (number of errors). Differences between groups also emerged for verbal IQ ($F(2,66) = 17.3, p < .05, MSE = 156.09$): LBs had less verbal knowledge than either EBs or monolinguals ($p < .05$ and $p < .001$, respectively), and EBs had less verbal knowledge than monolinguals ($p < .05$) (Bonferroni post hoc). There were also differences in verbal fluency ($F(2,66) = 5.8, p < .05, MSE = 0.69$): Bonferroni post hoc tests showed that both LBs and EBs performed worse than monolinguals ($p < .05$ and $p < .05$, respectively). No significant differences emerged between LBs and EBs.

Finally, results did not reveal any significant difference between the groups in either the decoding of words and pseudowords or nonverbal IQ. The absence of differences in the decoding of words and pseudowords, as measured by standardized tests, showed that the bilingual children were not characterized by reading difficulties or delays compared to monolinguals. Moreover, consistent with our selection criterion, all the children (bilingual and monolingual) were typically developing readers, as shown by the fact that none of their scores were more than two standard deviations below the mean, which can be regarded as the cut-off score for defining reading disorders.

Experiment 1: Lexicality and frequency effects

Experiment 1 investigated the lexicality effect, which refers to the advantage, in terms of error rates and reading speed, of words over pseudowords. Contrary to the claim that word reading is accomplished mainly via the nonlexical route in transparent orthographies, several studies have reported that Italian developing readers exhibit lexicality and frequency effects (e.g. Barca, Burani, Di Filippo & Zoccolotti, 2006; Paizi et al., 2011, 2013; Sulpizio & Colombo, 2013). These results indicate that lexical information is employed in reading

Italian from the very early years of learning to read. As to whether children who learn Italian as an L2 display similar lexical effects on their reading, we assumed that this is indeed the case, especially for EBs. Moreover, we predicted that EBs, characterized by a larger vocabulary than LBs (see [Table 1](#)), would exhibit the lexicality effect when low-frequency words were compared with matched pseudowords. By contrast, we expected that in the LB group the lexicality effect would be present for high-frequency words but not for low-frequency words, resulting in similar reading performances with respect to matched pseudowords.

Materials

We used a list of 60 stimuli. The experimental items were selected from the list originally developed by Pagliuca et al. (2008) for Italian adults reading aloud. Frequency values were taken from the CoLFIS frequency count for contemporary written Italian by Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando & Thornton (2005). All the stimuli were disyllabic, contained four to six letters, and were stressed on the penultimate syllable. The list of stimuli can be found in Appendix 1 and 2 (see also Primativo, O'Brien, Paizi, Rinaldi, Arduino & Burani, 2013).

Two dimensions were orthogonally manipulated: (i) regarding stimulus lexicality, half the stimuli were words (all nouns), whereas the other half were nonwords, matched with words on several psycholinguistic variables; (ii) regarding word frequency, half the word stimuli have a high frequency (HF) of occurrence in written Italian, and half have a low frequency (LF). The word stimuli were selected from the LEXVAR database (Barca, Burani & Arduino, 2002). Half the nonword stimuli were derived from the HF words and half from the LF words, by changing one or two letters. The final list included four sets of 15 items each, totaling 60 stimuli. Stimuli were matched for two initial phonemes (Kessler, Treiman & Mullennix, 2002), length in letters, bigram frequency, number of orthographic neighbors (N-size), summed neighbor frequency (Wagenmakers & Raaijmakers, 2006), and number of geminate consonants and diphthongs (Burani & Cafiero, 1991).

Procedure

The stimuli were presented in two separate blocks, of 30 trials each. Each block had a similar number of HF and LF words and pseudowords. The order of trials within each block was randomized, and block order was counterbalanced across participants. There was a practice block of six stimuli (three words and three pseudowords).

The stimuli were presented using E-Prime Version 2 (Psychology Software Tools, Pittsburgh, PA, USA;

<http://www.pstnet.com>) on a laptop computer (TravelMate 4000WLMi; Acer). Participants were asked to read aloud, as fast and accurately as possible, the stimuli that appeared in the center of the computer screen. Before each stimulus, an auditory cue was played for 250 ms, after which a fixation cross was displayed in the center of the screen for 250 ms. The stimulus appeared immediately after the fixation cross. Each stimulus disappeared as soon as the participant responded or after 5000 ms. There was an interstimulus interval of 250 ms.

A native Italian speaker noted and recorded the errors, and a second experimenter subsequently checked the recorded errors. A microphone connected to the laptop recorded the pronunciation of all the stimuli. Vocal reaction times (VRTs) were measured using SayWhen (Jansen & Watter, 2008), a software system designed to detect speech onset latencies automatically and with a high degree of accuracy, and which flags up a subset of trials most likely to have mismeasured onsets for optional manual checking. It allows experimenters to implement a graphical user interface that greatly speeds up and facilitates the checking and correction of this flagged subset of trials. Thanks to the use of this hybrid mode (automatic and hand coding), it is possible to achieve a very high level of accuracy compared with other fully automated methods (i.e. voice keys) or fully hand-coded methods.

Results

In both the first and second experiments, errors and VRTs (in ms) were analyzed using the generalized linear mixed model (GLMM), treating participants and items as crossed random factors. We chose to run this robust analysis in order to avoid a potential lack of power of the by-participant and by-item analyses, and to control for the variability of items and participants (Baayen, Davidson & Bates, 2008; for a discussion on the usefulness of GLMMs for examining variability in cognitive strategies during childhood, see Dauvier, Chevalier & Blaye, 2012). Errors were analyzed by means of the logistic model and VRTs by means of the linear one, as the residuals were basically normally distributed. We stepped through a series of models, then selected the best one according to the Akaike information criterion (AIC)⁵ value. All the comparisons between the models were carried out with the likelihood ratio test (LRT) using restricted maximum likelihood (REML) estimation. Thus, the best model was

⁵ The AIC is a measure of the relative quality of a statistical model, for a given set of data. In general, the AIC is $2k - 2\ln(L)$ where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model. Given a set of candidate models for the data, the preferred model is the one with the lowest AIC value (Akaike, 1974).

Table 2. Experiment 1. Mean (standard deviation) error percentages and vocal reaction times (in ms) for each group, for words (high- vs. low-frequency) and pseudowords (derived from high- and low-frequency words).

Experiment 1	Monolinguals mean (SD)	Early bilinguals mean (SD)	Late bilinguals mean (SD)
HF words_error %	0.9 (2.3)	1.3 (13.7)	1.8 (4)
LF words_error %	1.1 (2.9)	0.0 (0.0)	4.4 (6.5)
HF pseudowords_error %	3.5 (5.4)	2.7 (4.2)	5.3 (10.5)
LF pseudowords_error %	3.7 (8.2)	2.7 (4.9)	6.7 (9.1)
HF words_VRTs (ms)	709 (96)	696 (94)	771 (97)
LF words_VRTs (ms)	737 (112)	712 (92)	823 (109)
HF pseudowords_VRTs (ms)	772 (133)	754 (111)	850 (128)
LF pseudowords_VRTs (ms)	770 (117)	761 (124)	861 (142)

HF = high frequency, LF = low frequency, VRT = vocal reaction time

the one with the lowest AIC value and for which the comparison with other models was statistically significant. For this reason we only report this comparison. With this approach we rejected models that did not explain more variance than those we accepted.

For a clearer presentation of the results, Table 2 provides the mean values and standard deviations of error percentages and VRTs for each group with regard to words (HF and LF) and pseudowords (derived from HF and LF words).

Statistical analyses were conducted using lme4 (Bates, Maechler & Dai, 2008) and languageR packages (Baayen, 2008).

Errors

In the analysis of errors, we first tested a model (M0) that included participants and items as random factors and intercept as a fixed factor. We then added Group (monolinguals, EBs, and LBs), Frequency (HF vs. LF), and Lexicality (words vs. pseudowords) as fixed factors (M1). In a third model (M2) we added the interactions between the fixed factors (Group \times Frequency, Frequency \times Lexicality, Group \times Lexicality), and in a fourth one (M3) we finally added the Group \times Frequency \times Lexicality interaction. AIC values for all four models showed that M1 was the best model (AIC = 882.86; $LRT_{M0-M1} = 23.16$, $df = 4$, $p < .001$) compared with M0 (AIC = 898.02), M2 (AIC = 886.49), and M3 (AIC = 886.21). As shown in Table 3 below, a clear lexicality effect emerged in M1 (errors $\sim 1 + (1 | \text{participants}) + (1 | \text{items}) + \text{Group} + \text{Frequency} + \text{Lexicality}$), meaning that words were read more accurately than pseudowords (% errors, words: mean = 1.6, $SE = 0.3$; pseudowords: mean = 4.1, $SE = 0.9$). With regard to the group factor, results showed that LBs were less accurate (LBs: mean = 4.6, $SE = 0.8$; EBs: mean = 1.6, $SE = 0.5$; monolinguals:

mean = 2.3, $SE = 0.3$) than the other groups. The statistical parameters for all the factors are set out in Table 3.

Table 2 shows that there was a floor effect for EBs for LF words. This result was not problematic, as the GLMM with the logistic link function allows data to be characterized by an absence of variation, as the homogeneity of variance does not need to be satisfied.

In order to investigate the frequency effect, we ran a separate analysis on words only. In the selection of models, we proceeded as before. Again, we first tested a model (M0words) that included participants and items as random factors and intercept as a fixed factor. We then added Group (monolinguals, EBs, and LBs) and Frequency (HF vs. LF) as fixed factors (M1words). In a third model (M2words), we added the interaction between the two fixed factors (Group \times Frequency). AIC values for all four models showed that M1words was the best model (AIC = 262.23; $LRT_{M0words-M1words} = 8.99$, $df = 3$, $p < .05$) compared with M0words (AIC = 265.23) and M2words (AIC = 260.42). A group effect emerged in this model (errors $\sim 1 + (1 | \text{participants}) + (1 | \text{items}) + \text{Group} + \text{Frequency}$). The mean percentages of errors were 3.1% ($SE = 0.6$) for LBs, 0.7% ($SE = 0.6$) for EBs and 1% ($SE = 0.4$) for monolinguals. However, no frequency effect emerged. The statistical parameters are presented in Table 3.

Vocal reaction times

As with errors, we first tested a model (M0) that included participants and items as random factors and intercept as a fixed factor. We then added Group (monolinguals, EBs, and LBs), Frequency (HF vs. LF), and Lexicality (words vs. pseudowords) as fixed factors (M1). In a third model (M2), we added the interactions between the fixed factors (Group \times Frequency, Frequency \times Lexicality, Group \times Lexicality), and in a fourth one (M3) we finally added

Table 3. *Experiment 1. Statistical parameters of the generalized linear mixed models for errors and vocal reactions times.*

Model	Fixed factors	Estimate	SE	z value	p value
M1	Intercept	-4.2	0.3	-13.2	< .001
	group (EB)	-0.1	0.4	-0.3	ns
	group (LB)	0.9	0.4	2.2	< .05
	frequency (low)	0.2	0.3	0.9	ns
	Errors	lexicity (pseudowords)	1.2	0.3	4.1
M1words	Intercept	-6.3	0.6	-9.7	< .001
	group (EB)	-0.6	1.0	-0.6	ns
	group (LB)	1.5	0.6	2.3	< .05
	frequency (low)	0.6	0.6	1.0	ns
Model	Fixed factors	Estimate	SE	t value	p value
M3	Intercept	702.3	19.9	35.2	< .001
	group (EB)	-11.6	31.3	-0.4	ns
	group (LB)	61.0	31.3	1.9	< .05
	frequency (low)	28.1	16.5	1.7	< .05
	lexicity (pseudowords)	59.2	16.5	3.6	< .001
	group (EB) × frequency (low)	-8.8	13.13	-0.7	ns
	group (LB) × frequency (low)	17.9	13.3	1.3	ns
	group (EB) × lexicity (pseudowords)	-12.9	13.3	-1.0	ns
	group (LB) × lexicity (pseudowords)	2.0	13.4	0.15	ns
	frequency (low) × lexicity (pseudowords)	-27.2	23.3	-1.2	ns
	group (EB) × frequency (low) × lexicity (pseudowords)	13.7	18.7	0.7	ns
group (LB) × frequency (low) × lexicity (pseudowords)	-19.5	19.1	-1.0	ns	
M2words	Intercept	702.1	17.8	39.5	< .001
	group (EB)	-11.5	29.1	-0.4	ns
	group (LB)	61.1	29.1	2.1	< .05
	frequency (low)	28.3	13.5	2.1	< .05
	group (EB) × frequency (low)	-9.0	12.7	-0.7	ns
	group (LB) × frequency (low)	17.6	12.9	1.4	ns

EB = early bilinguals, LB = late bilinguals, VRT = vocal reaction time

the Group × Frequency × Lexicality interaction. AIC values for all four models showed that M3 was the best model (AIC = 47119.15; $LRT_{M2-M3} = 17.54$, $df = 2$, $p < .001$) compared with M0 (AIC = 47203.18), M1 (AIC = 47159.35), and M2 (AIC = 47132.69). A lexicity effect emerged in M3 (VRTs $\sim 1 + (1 | \text{participants}) + (1 | \text{items}) + \text{Group} + \text{Frequency} + \text{Lexicality} + \text{Group} : \text{Frequency} + \text{Frequency} : \text{Lexicality} + \text{Group} : \text{Lexicality} + \text{Group} : \text{Frequency} : \text{Lexicality}$), meaning that words (mean = 741, $SE = 13$) were read faster than pseudowords (mean = 795, $SE = 16$). Moreover, results showed group and frequency effects. The statistical parameters for all the factors are set out in Table 3.

In order to investigate the frequency effect, we ran a separate analysis on words only. We applied the same procedure as above. Three models were developed (M0words, M1words, and M2words). AIC values showed

that M2words was the best model (AIC = 23960.49; $LRT_{M1-M2} = 16.98$, $df = 2$, $p < .001$) compared with M0words (AIC = 24004.27) and M1words (AIC = 23973.47). With regard to M2words (VRTs $\sim 1 + (1 | \text{participants}) + (1 | \text{items}) + \text{Group} + \text{Frequency} + \text{Group} : \text{Frequency}$), a group and a frequency effect emerged. As before, the statistical parameters for this model are reported in Table 3 below.

Discussion

The results of the first experiment showed the presence of lexicity and frequency effects for all three groups of participants. These results confirmed that lexical information is employed in reading Italian even in the initial years of learning to read and, more importantly, they showed that this is also true for children who are

learning to read Italian as an L2, regardless of their AoE-L2. However, contrary to our expectations, the lexicality effect was not modulated by word frequency or by type of bilingualism (early vs. late), as there was no significant triple Group \times Lexicality \times Frequency interaction. Moreover, a group effect emerged, showing that LBs were less accurate and slower than monolinguals and EBs.

Experiment 2: Stress assignment

Experiment 2 was designed to explore the impact of lexicon size and AoE-L2 on reading Italian words with dominant and non-dominant stress. We also set out to investigate whether errors in reading are influenced by the distributional properties of the stress patterns in Italian. As is well known, although Italian has almost perfect grapheme-to-phoneme correspondence at a segmental level, there is a degree of unpredictability at the suprasegmental level, which can be regarded as a source of irregularity in this language. This is because the position of the stress in words with three or more syllables is neither orthographically marked nor predicted by rules. Most three- and four-syllable Italian words are stressed on the penultimate syllable, that is, they carry a dominant stress (e.g. *faRIna* “flour”). However, a smaller proportion of polysyllabic words are stressed on the antepenultimate syllable, that is, they carry a non-dominant stress (e.g. *FAvola* “tale”). Some authors have hypothesized that stress assignment is affected by stress neighborhood (i.e. the number of words that share the same stress pattern and final orthographic/phonological sequence) (Burani & Arduino, 2004; Burani et al., 2014; Colombo, 1992; Sulpizio, Arduino, Paizi & Burani, 2013). It has also been hypothesized that younger readers display greater sensitivity to stress dominance in reading because they have a smaller vocabulary than adult readers. The absence of lexical information about a word’s stress position may favor stress assignment to the dominant position, that is, the penultimate syllable (Paizi et al., 2011; Sulpizio et al., 2012; Sulpizio & Colombo, 2013).

In line with these considerations, assuming that bilingual readers (especially LBs with a limited lexicon) tend to overgeneralize stress assignment on the basis of stress dominance in L2, we would expect them to assign the most frequent stress pattern to the majority of stimuli (particularly low-frequency words).

Materials

The stimuli were adapted from the original list created by Paizi et al. (2011). The adapted list comprised four sets of three- and four-syllable words totaling 60 stimuli. There were 30 HF words and 30 LF words. Frequency was based on printed frequency counts for children (Marconi,

Ott, Pesenti, Ratti & Tavella, 1993). Half the words in each frequency set had the dominant stress pattern (i.e. on the penultimate syllable) and half the non-dominant stress pattern (i.e. on the antepenultimate syllable). Words in each subset (frequency–stress pattern) were matched for age of acquisition⁶ (AoA; Juhasz, 2005), familiarity, imageability, orthographic neighborhood size, length (in letters and syllables), bigram frequency, orthographic complexity, and initial phoneme. The number of words with the same final sequence and stress pattern (defined as STRESS FRIENDS; see Burani & Arduino, 2004; Burani et al., 2014) was balanced across the four sets of words.

The list of stimuli is provided in Appendix 1 and 2 (see also Primativo et al., 2013).

Procedure

The procedure was the same as in Experiment 1. Participants were asked to read the stimuli aloud as rapidly and accurately as possible. There was a practice block of four stimuli, two words with the dominant stress pattern and two with the non-dominant stress pattern. Stimuli were presented in random order, in two separate blocks in which stimuli were balanced for frequency (HF vs. LF) and stress pattern (penultimate vs. antepenultimate syllable). Block order was counterbalanced across participants.

As in Experiment 1, a native Italian speaker noted and recorded the errors, which were subsequently checked by a second experimenter. Owing to the ease with which word stress can be identified for each native Italian speaker, the stress errors were also noted by the experimenter on the fly. A microphone connected to the laptop recorded the pronunciation of all the stimuli. VRTs were measured by means of the SayWhen software system (Jansen & Watter, 2008).

The order in which Experiments 1 and 2 were administered was counterbalanced across participants.

Results

The mean percentages of pronunciation and stress errors for each group of participating readers, together with their VRTs (in ms), are presented in Table 4. The total percentage of errors was divided into percentages of pronunciation and stress assignment errors. Errors were classified as pronunciation errors when the participant did not accurately pronounce the word at the segmental level, that is, when the errors consisted of phoneme substitutions, omissions, insertions or transpositions, hesitations, stuttering or false starts. When the error consisted of incorrect stress placement, the response was

⁶ Here, age of acquisition refers to the age when a new word first enters the child’s lexicon.

Table 4. *Experiment 2. Mean (standard deviation) percentages of pronunciation and stress errors and vocal reaction times (in ms) for each group for words (high- vs. low-frequency) with dominant or non-dominant stress.*

Experiment 2	Monolinguals mean (SD)	Early bilinguals mean (SD)	Late bilinguals mean (SD)
HD words_P error %	0.6 (1.9)	0.0 (0.0)	3.1 (5.6)
LD words_P error %	1.3 (3.5)	0.9 (2.4)	4.9 (10.5)
HND words_P error %	1.3 (2.7)	0.0 (0.0)	1.3 (2.8)
LND words_P error %	0.9 (2.3)	1.3 (2.8)	4.0 (4.9)
HD words_S error %	0.4 (1.6)	1.3 (2.8)	0.4 (1.7)
LD words_S error %	2.0 (3.8)	3.1 (4.3)	2.2 (3.3)
HND words_S error S %	0.6 (1.9)	0.9 (2.4)	2.2 (4.1)
LND words_S error %	4.4 (5.3)	6.2 (4.7)	12.9 (12.2)
HD words_VRTs (ms)	727 (107)	744 (89)	878 (160)
LD words_VRTs (ms)	791 (129)	817 (139)	947 (205)
HND words_VRTs (ms)	735 (114)	770 (113)	868 (155)
LND words_VRTs (ms)	791 (144)	816 (129)	937 (185)

HD = high frequency dominant, LD = low frequency dominant, HND = high frequency non-dominant, LND = low frequency nondominant, S = stress, P = pronunciation, VRT = vocal reaction time

classified as a stress error. If a pronunciation and a stress error were both present in the response, the error was deemed to be a pronunciation error.

As in Experiment 1, errors (separated into pronunciation and stress errors) and VRTs were analyzed using the GLMM, treating participants and items as crossed random factors. More specifically, errors were analyzed with the logistic model and VRTs with the linear one as the residuals were basically normally distributed. Once again, we explored a series of models, then we selected the best one according to its AIC value. As before, all the comparisons between the models were carried out with the LRT, using REML estimation. As in Experiment 1 (see Experiment 1, Results section) we rejected models that did not explain more variance than those we accepted, and only the results for the best model are reported. Statistical analyses were conducted using the lme4 (Bates et al., 2008) and languageR packages (Baayen, 2008).

Errors

In the analysis of pronunciation errors, we first tested a model (M0) that included participants and items as random factors. In a second model (M1), we added Group (monolinguals, EBs, and LBs), Stress (dominant vs. non-dominant) and Frequency (HF vs. LF) as fixed factors. At a later stage, we added the interactions between the fixed factors (Group \times Frequency, Frequency \times Stress, Group \times Stress) (M2), and finally added the Group \times Frequency \times Stress interaction (M3). AIC values for these

four models showed that M1 was the best model (AIC = 563.01; $LRT_{M0words-M1words} = 24.40$, $df = 4$, $p < .001$) compared with M0 (AIC = 581.41), M2 (AIC = 568.73) and M3 (AIC = 571.12). M1 (pronunciation errors $\sim 1 + (1 | participants) + (1 | items) + Group + Stress + Frequency$) showed a group effect (% error: monolinguals: mean = 1, $SE = 0.5$; EBs: mean = 0.6, $SE = 0.7$; LBs: mean = 3.3, $SE = 0.7$). It also revealed an effect of frequency, suggesting that LF items (mean = 2.2, $SE = 0.5$) were read less correctly than HF ones (mean = 1.1, $SE = 0.3$). Neither the main effect of stress pattern nor the interaction effects were significant. Statistical parameters for this model are shown in Table 5.

Table 4 shows that there was a floor effect for early EBs for HF words with dominant and non-dominant stress. Once again, this result was not problematic as the GLMM with a logistic link function allows data characterized by an absence of variation to be processed because the homogeneity of variance does not need to be satisfied.

With regard to the analysis of stress errors, as we had a clearly defined experimental hypothesis concerning this variable (LBs assign the most frequent stress pattern to LF words), we created a new independent variable (latebil_lowfreq_non-domistress) that allowed us to distinguish items belonging to these three conditions (i.e. answered by LBs, non-dominant stress, LF). This procedure followed the one commonly used with planned contrasts in linear models (for a comprehensive explanation, see Noel, 2013). Moreover, by using this

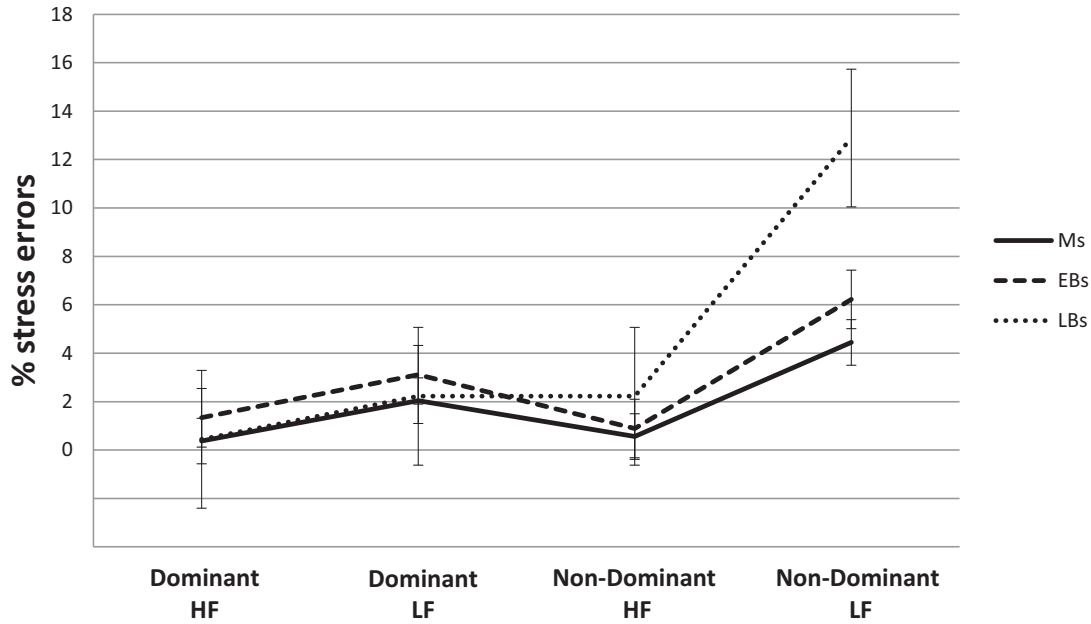


Figure 1. Percentages of stress assignment errors as a function of type of stress pattern (Dominant/Non-dominant), word frequency (High and Low) and group (monolinguals, early bilinguals and late bilinguals) (error bars show standard error). Ms = monolinguals, EBs = early bilinguals, LBs = late bilinguals, HF = high frequency, LF = low frequency

specific variable, we were able to take into account the smallest number of parameters (principle of parsimony), which is in line with the use of AIC values to select the best fitting model.

We first tested a model (M0) that included participants and items as random factors. In a second model (M1), we added Group (monolinguals, EBs, and LBs), Stress (dominant vs. non-dominant) and Frequency (HF vs. LF) as fixed factors. At a later stage, we added the interactions between these fixed factors (Group × Frequency, Frequency × Stress, Group × Stress) (M2), and subsequently added the Group × Frequency × Lexicality interaction (M3). Finally, we considered a fifth model (M4) that included participants and items as random factors, and group, stress, frequency, and *latebil_lowfreq_non-domistress* as fixed factors. AIC values clearly showed that M4 was the best model (AIC = 710.10; $LRT_{M1-M4} = 6.88$, $df = 1$, $p < .05$) compared with M0 (AIC = 735.46), M1 (AIC = 716.19), M2 (AIC = 715.79), and M3 (AIC = 719.37). In particular, M4 (stress errors $\sim 1 + (1 | participants) + (1 | items) + Group + Frequency + Stress + latebil_lowfreq_non-domistress$) showed an effect of frequency, with LF words being read less accurately than HF words (LF items: mean = 5.2, $SE = 0.6$; HF items: mean: 1, $SE = 0.3$). Moreover, we observed an effect of the *latebil_lowfreq_non-domistress* factor, suggesting that LBs were less accurate than EBs and monolinguals on LF words with non-dominant stress (see Figure 1). The statistical parameters are presented in Table 5.

Vocal reaction times

Once again, we began by testing a model (M0) that included participants and items as random factors and intercept as a fixed factor. We then added Group (monolinguals, EBs, and LBs), Frequency (HF vs. LF), and Stress (dominant vs. non-dominant) as fixed factors (M1). In a third model (M2), we added the interactions between the fixed factors (Group × Frequency, Frequency × Stress, Group × Stress), and in a fourth one (M3) we finally added the Group × Frequency × Stress interaction. AIC values for these four models showed that M3 was the best model (AIC = 47717.34; $LRT_{M2-M3} = 18.05$, $df = 2$, $p < .001$) compared with M0 (AIC = 47822.38), M1 (AIC = 47760.16), and M2 (AIC = 47731.39). In M3 (VRTs $\sim 1 + (1 | participants) + (1 | items) + Group + Frequency + Stress + Group : Frequency + Frequency : Stress + Group : Stress + Group : Frequency : Stress$), a frequency effect emerged, meaning that HF words were read faster than LF ones. There was also an effect of group, but neither the main effect of stress pattern nor the interaction effects were significant. The statistical parameters for all the factors are reported in Table 5 below.

Discussion

To sum up, the results of Experiment 2 highlighted the presence of a frequency effect in pronunciation errors and stress errors, both for monolinguals and bilinguals, regardless of AoE-L2. Percentages of pronunciation errors showed a group effect, indicating that LBs were less

Table 5. *Experiment 2. Statistical parameters of the generalized linear mixed models for errors and vocal reactions times.*

		Model	Fixed factors	Estimate	SE	z value	p value	
Pronunciation	M1	Intercept		-5.7	0.4	-13.5	< .001	
		group (EB)		-0.6	0.7	-0.8	ns	
		group (LB)		1.8	0.4	4.2	< .001	
		frequency (low)		0.7	0.3	2.2	< .05	
		stress (non-dominant)		0.1	0.3	0.2	ns	
Errors	Stress	M4	Intercept		-7.0	0.8	-9.1	< .001
			group (EB)		0.7	0.4	1.7	ns
			group (LB)		0.6	0.5	1.3	ns
			frequency (low)		2.1	0.8	2.6	< .05
			stress (non-dominant)		0.9	0.7	1.2	ns
			latebil_lowfreq_non-domistress		1.3	0.5	2.5	< .05
		Model	Fixed factors	Estimate	SE	t value	p value	
VRTs	M3	Intercept		730.7	24.5	29.8	< .001	
		group (EB)		9.5	39.2	0.2	ns	
		group (LB)		127.6	39.2	3.2	< .001	
		frequency (low)		70.8	19.4	3.6	< .001	
		stress (non-dominant)		9.2	19.4	0.5	ns.	
		group (EB) × frequency (low)		15.5	16.4	0.9	ns	
		group (LB) × frequency (low)		22.7	16.6	1.3	ns	
		group (EB) × stress (non-dominant)		22.05	16.3	1.3	ns	
		group (LB) × stress (non-dominant)		-2.0	16.5	-0.1	ns	
		frequency (low) × stress (non-dominant)		-0.4	27.5	-0.0	ns	
		group (EB) × frequency (low) × stress (non-dominant)		-28.5	23.3	-1.2	ns	
group (LB) × frequency (low) × stress (non-dominant)		-21.3	23.8	-0.9	ns			

EB = early bilinguals, LB = late bilinguals, VRT = vocal reaction time

accurate than EBs and monolinguals. However, there was no stress pattern effect and no interactions between stress pattern, frequency and group. The most interesting result was for stress assignment errors, as we found that LBs were less accurate than EBs and monolinguals on LF words with a non-dominant (antepenultimate syllable) stress. This result is discussed below.

Finally, the groups differed on VRTs, with LBs being the slowest group. However, VRTs (for correctly read items) were not sensitive to stress patterns, confirming similar findings for both adults and children (see, for example, Burani et al., 2013; Primativo et al., 2013; Sulpizio et al., 2012). We therefore do not discuss this result in the General Discussion.

General discussion

The purpose of the present study was twofold. First, we wanted to investigate lexicality and frequency effects in

bilingual children reading Italian as an L2 and differing in AoE-L2. Second, we wanted to examine suprasegmental processing (i.e. stress assignment) in reading Italian as an L2 in the same groups of children. The role of AoE-L2 was studied by comparing two groups of bilingual children: one characterized by an AoE-L2 before three years and eleven months old (early bilinguals) and one characterized by an AoE-L2 above four years (late bilinguals).

The results of Experiment 1 showed that words were read better than pseudowords, that is, a lexicality effect emerged in all three groups. In addition, high-frequency words were read faster than low-frequency words. Thus, similarly to monolinguals, bilingual children exhibited sensitivity to the lexical properties of Italian words. As for monolinguals, results confirmed those of previous studies highlighting the importance of a lexical strategy in reading Italian (e.g. Barca et al., 2006; Paizi et al., 2011; 2013). Lexical information is employed when reading Italian as

an L2 in the same way as it is when reading Italian as an L1. We observed that late bilinguals were less accurate than both monolinguals and early bilinguals. However, both groups of bilinguals exhibited a similar lexicality effect to monolinguals, reading words better than pseudowords. The absence of a significant interaction between group and lexicality invalidated our hypothesis of a greater lexicality effect for early bilinguals than for late bilinguals. However, the absence of a significant interaction does not mean that the groups had similar lexical knowledge. An inspection of the mean VRTs and error percentages for participants in Experiment 1 (Table 2) reveals that late bilinguals exhibited greater differences between high-frequency and low-frequency words (on both VRTs and errors) than the other groups, although these differences did not reach statistical significance. Larger frequency effects in younger or less skilled readers have often been reported (for Italian, see, for example, Barca et al., 2006; Mazzotta, Barca, Marcolini, Stella & Burani, 2005), and have been attributed to the fact that several low-frequency words are likely to be unknown to readers with a limited vocabulary. The tendency toward a larger frequency effect for late bilinguals is thus consistent with the idea of a limited vocabulary, especially in the case of low-frequency words. Then again, it may be that testing high-frequency versus low-frequency words is not enough to reveal differences between late bilinguals' and early bilinguals' (or monolinguals') lexical reading: other word properties (e.g. prosodic properties such as lexical stress) may be more revealing than a mere high-frequency versus low-frequency word comparison. This point deserves to be explored further in future research. In general, this pattern of results supports the hypothesis of overall lexical influence in reading in transparent orthographies. To the best of our knowledge, this is the first time that this result has been found for L2 Italian, and it is consistent with similar findings for L1 and L2 English (e.g. Erdos, Genesee, Savage & Haigh, 2010; Erdos, Genesee, Savage & Haigh, 2014; Genesee, 2007; Genesee & Lindholm-Leary, 2012).

The results of Experiment 2 indicated that low-frequency words with non-dominant stress received the correct stress less frequently than dominant-stress words. Critically to the purpose of the present study, different patterns were obtained with regard to AoE-L2: only late bilinguals were significantly deficient in reading low-frequency words with the non-dominant antepenultimate stress pattern. These results support the idea that less skilled readers may exhibit greater sensitivity to stress dominance in reading, possibly owing to their limited lexical knowledge compared with that of skilled readers. An absence of lexical information about stress position, with a consequent reliance on sublexical processing (Ziegler & Goswami, 2005), may lead children to apply the statistically dominant stress, thus

assigning the dominant/most frequent stress pattern to less familiar words. This finding was reported, by Sulpizio and Colombo (2013), in second graders and, to a lesser extent, fourth graders reading low-frequency words, and was also observed in a group of older children with developmental dyslexia by Paizi et al. (2011). In a similar vein, Sulpizio et al. (2012) found that when they read nonwords aloud, second graders assigned the dominant stress more often than the non-dominant stress – a tendency that disappeared with age (for similar results in English, see Arciuli, Monaghan & Seva, 2010). Accordingly, readers with a reduced lexicon may tend to make more errors when reading unfamiliar words with a non-dominant stress. If we look at the results of the screening tests (Table 1) we can see that late bilinguals had a significantly lower verbal IQ than early bilinguals and monolinguals. However, neither late bilinguals nor early bilinguals were impaired in word and pseudoword reading, as assessed by means of a standardized test. In other words, none of their scores in the standardized tests were more than two standard deviations below the mean, which is regarded as the cut-off score for defining reading disability. This means that our late bilinguals group was not composed of poor readers, consistent with our selection criterion. This allows us to assume that late bilinguals had developed adequate decoding strategies, even though they scored lower than the other two groups, particularly when reading low-frequency words with a non-dominant stress. In this specific condition, vocabulary size may have had a greater impact on stress assignment. As Sulpizio and Colombo (2013) have suggested, in younger (Grade Two) readers, lexical reading is still under development and the lexicon still limited, with incomplete knowledge of word stress. Accordingly, late bilinguals (like younger and less skilled Italian readers) made more errors when reading unfamiliar L2 words with a non-dominant stress because of their limited L2 vocabulary, not because of a general reading difficulty. It also has to be underscored that the results on standardized measures showed that early bilinguals performed at the same level as – and sometimes better than – monolinguals, providing added evidence that early bilingualism does not delay linguistic development and literacy acquisition (e.g. Paradis et al., 2011).

With regard to the type of lexicon that may influence stress assignment in less skilled readers, Sulpizio and Colombo (2013) have suggested that at least the orthographic one might be involved. We did not directly address this question in the present study where lexicon size was measured by means of two tasks: receptive vocabulary and verbal fluency. That is, lexicon size here essentially meant oral vocabulary size, as is usual in the literature on bilingualism. We did not record possible stress errors during oral production, which would have allowed us to discuss the potential influence of the

phonological lexicon on stress assignment in reading. In principle, late bilinguals can be assumed to have both a limited phonological and a limited orthographic lexicon. Thus, they would have difficulty rapidly recognizing printed low-frequency words, with consequent difficulty matching orthographic and phonological representations. This is an interesting issue that needs to be explored in future research.

To sum up, the present study showed that bilingual children learning to read in an L2 are sensitive to the distributional properties of the target L2 language. Specifically, in the case of a language with a transparent orthography such as Italian, we found that the performance of bilingual children in a reading aloud task was affected by lexicality (they read words better than pseudowords), word frequency (they read high-frequency words better than low-frequency words), and stress pattern (they tended to assign the dominant stress to low-frequency words with a non-dominant stress). Furthermore, these effects were modulated by AoE-L2, in that late bilinguals performed worse than early bilinguals when it came to assigning stress to low-frequency words – an effect we attribute to their reduced vocabulary size.

One possible objection to this interpretation is that lexicon size and AoE-L2 cannot be easily disentangled. Thus, the effects found for early bilinguals and late bilinguals are intertwined with differences in lexicon size. Despite the debate in the literature regarding the presence of a critical period for L2 acquisition, it is an ecological fact that developing readers have different lexicon sizes according to AoE-L2. That is, the results of the present study could be interpreted in terms of AGE EFFECTS on vocabulary size and subsequent lexical reading development. This is a broader concept that also considers the AMOUNT of exposure to L2, which can affect all aspects of proficiency (e.g. vocabulary). Besides the amount of exposure, other factors can influence proficiency, such as the TYPE of exposure children have to L2 (for a comprehensive review, see Paradis et al., 2011). However, these issues are beyond the ambit of our study.

The findings of the present study lead to a further consideration: despite their poor vocabulary, late bilinguals develop a high degree of sensitivity to the systematic linguistic properties of their L2. The literature reports contrasting results on the bilingual advantage in terms of metalinguistic awareness. In a recent study, Bialystok and Barac (2012, p. 72) advanced the hypothesis that “bilingualism gives children a boost into figuring out structural relations within language but beyond that initial insight, more bilingual experience does not move that development forward”. The authors observed that metalinguistic awareness improved as a function of

language proficiency, but not bilingual exposure. We did not test metalinguistic awareness. In further research, it might be worthwhile looking at whether sensitivity to the lexical properties of the L2 and reading skills are related to metalinguistic awareness in early bilinguals and late bilinguals. Finally, it seems important to point out that both early bilinguals and late bilinguals displayed good skills in using the sublexical route for reading (they read nonwords as accurately as monolinguals did). From a clinical standpoint, this means that a behavioral delay in reading accuracy for late bilinguals does not necessarily reflect the presence of a specific reading disorder in L2. In most cases, this delay could be ascribed to poor vocabulary size, which reduces word reading speed. By contrast, there is a broad consensus that the most sensitive measure for detecting specific reading disorders in L2 is the reading speed for nonwords (e.g. Erdos et al., 2014), particularly for transparent languages.

To conclude, we argue that investigating how bilingual children learn to read increases our knowledge of the processes involved in learning to read in L1. Further research is needed to define precisely the nature of reading development in L2, and studies involving L1s and L2s with different orthographic structures (transparent vs. opaque) should be encouraged.

Appendix 1. Stimuli used in Experiment 1

HF words	LF words	HF pseudowords *	LF pseudowords *
barba	belva	barta	bemo
fiume	fieno	fiuro	fiemi
fuoco	fune	farto	futa
fretta	frate	friste	frala
latte	lite	liere	lipe
luna	lutto	lafo	litta
mondo	muffa	mopra	musta
neve	nuora	namo	nuoso
pane	pala	pono	peba
pietra	piuma	piese	piura
ponte	panno	pavo	padda
riva	rata	rimi	ranno
sonno	sorso	soppo	sorra
testa	tordo	trillo	torla
vino	vetta	vima	venso

HF = high frequency, LF = low frequency

* Pseudowords derived from the HF and LF words, respectively.

Appendix 2. Words used in Experiment 2

Dominant HF words	Dominant LF words	Non-dominant HF words	Non-dominant LF words
alunno	bistecca	angolo	bibita
bastone	cantiere	carcere	canapa
cervello	catino	diavolo	cocomero
coniglio	erede	favola	edicola
farina	fanale	femmina	forfora
fucile	furgone	nuvola	fulmine
moneta	municipio	ospite	incudine
odore	padella	pagina	lapide
palazzo	parrucca	pecora	pettine
parete	pattuglia	pentola	pollice
parola	pomata	pericolo	pontefice
piscina	scaffale	polvere	pugile
stagione	tamburo	scatola	scapolo
tappeto	uragano	tavolo	tenebra
tartaruga	zanzara	zucchero	zingaro

HF = high frequency, LF = low frequency

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Language, Memory, and Cognition*, 18, 234–254.
- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology*, 59, 920–948.
- Arciuli, J., Monaghan, P., & Ševa, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, 63, 180–196.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Minority-Language Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, 34, 424–434.
- Barca, L., Burani, C., Di Filippo, G., & Zoccolotti, P. (2006). Italian developmental dyslexic and proficient readers: Where are the differences? *Brain and Language*, 98, 347–351.
- Bates, D. M., Maechler, M., & Dai, B. (2008). *lme4: Linear mixed-effects models using Eigen and S4 classes*. R package version 0.999375–24. [Retrieved from <http://CRAN.R-project.org/package=lme4>].
- Bellocchi, S. (2010). *Una proposta di questionario per la raccolta di informazioni sulla storia linguistica di bambini bilingui* [A proposal for a questionnaire to collect information on bilingual children's linguistic history]. Poster presented at the 19th national conference of AIRIPA, Ivrea, Italy.
- Bellocchi, S., & Bastien-Toniazzo, M. (2011). Normal reading abilities and specific reading disabilities (developmental dyslexia): A cross linguistic study. *Rivista di Psicologica Applicata*, 1–2, 69–86.
- Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)* [CoLFIS. Corpus and Frequency Lexicon of Written Italian]. <http://www.istc.cnr.it/grouppage/colfisEng>.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12, 3–11.
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122, 67–73.
- Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language*, 109, 93–100.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, pp. 162–181. Mahwah, NJ: Lawrence Erlbaum.
- Bisiacchi, P. S., Cendron, M., Gugliotta, M., Tressoldi, P. E., & Vio, C. (2005). *BVN 5–11 – Batteria di valutazione neuropsicologica per l'età evolutiva* [Test for neuropsychological assessment]. Trento: Erickson.
- Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science*, 14, 256–269.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, 90, 318–325.
- Burani, C., & Cafiero, R. (1991). The role of sub-syllabic structure in lexical access to printed words. *Psychological Research*, 53, 42–52.
- Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpho-lexical reading develop in readers of a shallow orthography? *Brain and Language*, 81, 568–586.
- Burani, C., Paizi, D., & Sulpizio, S. (2014). Stress assignment in Italian: Friendship outweighs dominance. *Memory & Cognition*, 42, 662–675.
- Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 987–1003.

- Colombo, L., Pasini, M., & Balota, A. D. (2006). Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance. *Memory & Cognition*, *34*, 1312–1324.
- Contento, S., Bellocchi, S., & Bonifacci, P. (2013). *BaBIL. Prove per la valutazione delle competenze verbali e non verbali in bambini bilingui* [BaBIL: Tests for the assessment of verbal and non-verbal abilities in bilingual children]. Florence: Giunti O.S.
- Cornoldi, C., & Colpo, G. (1998). *Prove di lettura MT per la scuola elementare* [MT Reading tests for primary schools]. Florence: Giunti O.S.
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86.
- Dauvier, B., Chevalier, N., & Blaye, A. (2012). Using finite mixture of GLMs to explore variability in children's flexibility in a task-switching paradigm. *Cognitive Development*, *27*, 440–454.
- Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A distressing “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*, 1606–1618.
- Erdos, C., Genesee, F., Savage, R., & Haigh, C. (2010). Individual differences in second language reading outcomes. *International Journal of Bilingualism*, *15*, 3–25.
- Erdos, C., Genesee, F., Savage, R., & Haigh, C. (2014). Predicting risk for oral and written language learning difficulties in students educated in a second language. *Applied Psycholinguistics*, *35*, 371–398.
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.
- Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, *16*, 161–179.
- Genesee, F. (2007). French immersion and at-risk students: A review of research findings. *Canadian Modern Language Review*, *63*, 655–688.
- Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology*, *49*, 140–147.
- Genesee, F., & Lindholm-Leary, K. (2012). The education of English language learners. In K. Harris, S. Graham & T. Urdan (eds.), *APA educational psychology handbook*, Washington, DC: APA Books, pp. 499–526.
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. New York: Cambridge University Press.
- Goetry, V., Wade-Woolley, L., Kolinsky, R., & Mousty, P. (2006). The role of stress processing abilities in the development of bilingual reading. *Journal of Research in Reading*, *29*, 349–362.
- Holowka, S., Brosseau-Lapré, F., & Petitto, L. A. (2002). Semantic and conceptual knowledge underlying bilingual babies' first signs and words. *Language Learning*, *52*, 205–262.
- Isel, F., Baumgärtner, A., Thraen, J., Meisel, J. M., & Buechel, C. (2010). Neural circuitry of the bilingual mental lexicon: Effect of age of second language acquisition. *Brain and Cognition*, *72*, 169–180.
- Jansen, P. A., & Watter, S. (2008). SayWhen: An automated method for high-accuracy speech onset detection. *Behavior Research Methods*, *40*, 744–751.
- Johnson, J. S., & Newport, E. L. (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.
- Juhász, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test*. Circle Pines, MN: AGS.
- Kelly, M. H., Morris, J., & Verrechia, L. (1998). Orthographic cues to lexical stress: Effects on naming and lexical decision. *Memory & Cognition*, *26*, 822–832.
- Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, *47*, 145–171.
- Kirby, J. R., Desrochers, A., Roth, L., & Sandy, S. V. L. (2008). Longitudinal predictors of word reading development. *Canadian Psychology*, *49*, 103–110.
- Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, *11*, 203–223.
- Krämer, M. (2009). *The phonology of Italian*. New York: Oxford University Press.
- Lenneberg, E. H. (1967). *Biological foundations of language*. Oxford: Wiley.
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language*, *58*, 265–326.
- Marconi, L., Ott, M., Pesenti, E., Ratti, D., & Tavella, M. (1993). *Lessico elementare. Dati statistici sull'italiano letto e scritto dai bambini delle elementari* [Elementary lexicon: Statistical data for Italian written and read by elementary school children]. Bologna: Zanichelli.
- Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*, 9–34.
- Mazzotta, S., Barca, L., Marcolini, S., Stella, G., & Burani, C. (2005). *Frequenza, immaginabilità ed età di acquisizione delle parole: in che misura influenzano la lettura dei bambini italiani?* [Frequency, imageability, and age-of-acquisition of words: How do they affect Italian children's reading aloud?] *Psicologia Clinica dello Sviluppo*, *9*, 247–266.
- Noel, Y. (2013). *Psychologie statistique avec R* [Statistic psychology with R]. Paris: Springer.

- Pagliuca, G., Arduino, L. S., Barca, L., & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: The lexicality effect in Italian. *Language and Cognitive Processes*, 23, 422–433.
- Paizi, D., De Luca, M., Zoccolotti, P., & Burani, C. (2013). A comprehensive evaluation of lexical reading in Italian developmental dyslexics. *Journal of Research in Reading*, 3, 303–329.
- Paizi, D., Zoccolotti, P., & Burani, C. (2011). Lexical stress assignment in Italian developmental dyslexia. *Reading and Writing*, 24, 443–461.
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders*. Baltimore, MD: Brookes Publishing.
- Pearson, B. Z. (1998). Assessing lexical development in bilingual babies and toddlers. *International Journal of Bilingualism*, 2, 347–372.
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93–120.
- Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.
- Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., & Cappa, S. F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19, 170–182.
- Petitto, L. A., Katerelos, M., Levy, B., Gauna, K., Tétrault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for mechanisms underlying bilingual language acquisition. *Journal of Child Language*, 28, 1–44.
- Petitto, L. A., & Kovelman, I. (2003). The bilingual paradox: How signing-speaking bilingual children help us resolve bilingual issues and teach us about the brain's mechanisms underlying all language acquisition. *Learning Languages*, 8, 5–18.
- Primativo, S., O'Brien, S., Paizi, D., Rinaldi, P., Arduino, L. S., & Burani, C. (2013). Bilingual vocabulary size and lexical reading in Italian. *Acta Psychologica*, 144, 554–562.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory & Language*, 42, 342–364.
- Sartori, G., Job, R., & Tressoldi, P. E. (2007). *DDE-2. Batteria per la valutazione della Dislessia e della Disortografia Evolutiva* [Developmental dyslexia and dysorthographia assessment battery]. Florence: Giunti O.S.
- Seidenberg, M. S., Waters, G. S., Barnes, M., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, 23, 383–404.
- Slavin, R. E., & Cheung, A. (2003). *Effective programs for English language learners: A best-evidence synthesis*. Baltimore, MD: CRESPAR.
- Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 51–68.
- Sulpizio, S., & Colombo, L. (2013). Lexical stress, frequency, and stress neighborhood effects in the early stages of Italian reading development. *Quarterly Journal of Experimental Psychology*, 66, 2073–2084.
- Sulpizio, S., Boureux, M., Burani, C., Deguchi, C., & Colombo, L. (2012). Stress assignment in the development of reading aloud: Nonword priming effects on Italian children. In N. Miyake, D. Peebles & R. P. Cooper (eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pp. 2369–2374. Austin, TX: Cognitive Science Society.
- Thornton, A. M., Iacobini, C., & Burani, C. (1997). *BDVDB. Una base di dati sul vocabolario di base della lingua italiana* [BDVDB: A database for the Italian basic dictionary]. Rome: Bulzoni.
- Uchikoshi, Y., & Marinova-Todd, S. H. (2012). Language proficiency and early literacy skills of Cantonese-speaking English language learners in the U.S. and Canada. *Reading and Writing*, 25, 2107–2129.
- Wagenmakers, E. J., & Raaijmakers, J. G. W. (2006). Long-term priming of neighbors biases the word recognition process: Evidence from a lexical decision task. *Canadian Journal of Experimental Psychology*, 60, 275–284.
- Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29.