READING DEVELOPMENT IN TWO ALPHABETIC SYSTEMS DIFFERING IN ORTHOGRAPHIC CONSISTENCY: A LONGITUDINAL STUDY OF FRENCH-SPEAKING CHILDREN ENROLLED IN A DUTCH IMMERSION PROGRAM

Katia LECOCQ, Régine KOLINSKY, Vincent GOETRY, José MORAIS,
Jesus ALEGRIA, & Philippe MOUSTY
Université Libre de Bruxelles

Studies examining reading development in bilinguals have led to conflicting conclusions regarding the language in which reading development should take place first. Whereas some studies suggest that reading instruction should take place in the most proficient language first, other studies suggest that reading acquisition should take place in the most consistent orthographic system first.

The present study examined two research questions: (1) the relative impact of oral proficiency and orthographic transparency in second-language reading acquisition, and (2) the influence of reading acquisition in one language on the development of reading skills in the other language.

To examine these questions, we compared reading development in French-native children attending a Dutch immersion program and learning to read either in Dutch first (most consistent orthography) or in French first (least consistent orthography but native language). Following a longitudinal design, the data were gathered over different sessions spanning from Grade 1 to Grade 3. The children in immersion were presented with a series of experimental and standardised tasks examining their levels of oral proficiency as well as their reading abilities in their first and, subsequently in their second, languages of reading instruction. Their performances were compared to the ones of French and Dutch monolinguals.

The results showed that by the end of Grade 2, the children instructed to read in Dutch first read in both languages as well as their monolingual peers. In contrast, the children instructed to read in French first lagged behind the other Dutch-speaking groups in Dutch reading tasks.

These findings extend the notion that differences across languages in terms of orthographic transparency impact on reading development to the French-Dutch pair, and strongly support the view that there are potentially significant benefits to learn to read in the most consistent orthographic system first, even though it is the least proficient language, since it boosts phonological processes and improves subsequent reading acquisition in both languages.

Introduction

With our modern means of communication and travelling, learning to read and write in a second language (henceforth, *L2*) is becoming the rule rather than the exception. In the last decades, L2 acquisition through *immersion* has gained widespread acceptance, after it was introduced in the 1960s in Canada and since then in the United States as well as in Belgium. Unlike traditional language courses in which the L2 is the subject material, language immersion is a method which uses the L2 as a teaching tool, surrounding or "immersing" students in that language through the teaching of various topics like mathematics, history, etc.

Extensive, systematic evaluations of immersion programs in a number of Canadian cities have provided strong evidence that these programs are remarkably effective. However, many practical questions remain unanswered regarding how to implement immersion programs. For example, at what age should the immersion begin? Is it disadvantageous for the development of the mother tongue or not? In which language should the children be taught to read and spell first? The present study focuses on this last question.

The role of oral proficiency in second-language reading acquisition

Despite 40 years of research on literacy acquisition in a L2, researchers still disagree on the issue of whether reading instruction should take place in the native language first (henceforth, *L1*), or not. Some studies suggest that reading acquisition should take place in the native language first because successful reading acquisition requires a minimum of oral proficiency. For example, Verhoeven (2000) compared the development of reading and spelling in Turkish-native children schooled in Dutch in the Netherlands and in Dutch-native monolingual children, in the first two grades of elementary school. This comparison showed that the Turkish-native children reached much lower performances than their monolingual peers.

To account for this difference, Verhoeven (2000) suggested that L2 learners experience difficulties in recoding letter strings phonemically because they are less able than native speakers to discriminate the sounds of that language. This hypothesis is supported by several studies showing that L2 learners have difficulties processing non-native phonemic contrasts. A well-

known example is the difficulty of Japanese-native speakers to distinguish between the phonemes /r/ and /l/ (Goto, 1971). More recently, Sundara, Polka, and Genesee (2006) compared 4-year-old monolingual (English or French) and bilingual children, as well as bilingual adults who had learned these two languages simultaneously, on their ability to discriminate the English contrast /d-ð/. They showed that although the ability to discriminate this contrast improved with age both in the English children and in the bilinguals, the discrimination abilities of French-speaking children and adults, who had no experience with this contrast, remained poor and unchanged during development. However, Sebastian-Gallés and Soto-Faraco (1999) showed that, in a task consisting in identifying phonemic contrasts which exist in Catalan but not in Spanish, even highly proficient Spanish-Catalan bilinguals who had acquired their L2 within their first years of life and used both languages to the same extent as adults, systematically performed worse, and needed longer portions of the signal, than Catalan-Spanish bilinguals. A study by Wade-Woolley and Geva (2000) also indicates that English-native speakers learning Hebrew experienced more difficulties in discriminating the phonemic contrast /ts/ vs. /s/ when it occurred in syllable onsets, which is found in Hebrew but not in English, than when it occurred in rimes, a structural context which is found in both languages. Moreover, accuracy on this measure was correlated to word reading abilities in both languages. This whole set of results across studies suggests that phonological elements specific to the L2 present additional challenges to beginning readers, and that even under conditions of early and extensive exposure, non-native phonemic categories are not processed with the same degree of efficiency as L1 categories.

According to Verhoeven (2000), L2 learners might also have difficulties with building up a reading vocabulary in their L2 because of the restricted size of their lexicon in that language, which may seriously impede reading in L2 both for words and for texts. Indeed, due to smaller vocabulary knowledge, L2 learners might benefit less from the effect of word frequency during lexical access. At the same time, their poor vocabulary knowledge might interfere with text comprehension. In Verhoeven's (2000) study, the Turkish-native children lagged more than 2 standard deviations behind their Dutch-native peers in vocabulary knowledge in Dutch. With regards to reading comprehension, these children also showed substantially lower levels of achievement than the Dutch monolinguals. These results suggest that children who learn to read in a non-proficient L2 first may have greater difficulties developing both the phonological and the lexical word recognition procedures than children who learn to read in their L1. Indeed, they may experience difficulties not only in recoding letter strings phonemically, but also in building up a reading vocabulary in their L2.

However, the children examined by Verhoeven (2000) came from ethnic minorities and did not benefit from any additional school or home support. Therefore, socio-economic factors may account for (at least part of) their poor performances compared to the Dutch monolinguals. Moreover, recent studies on word recognition in monolinguals reading in different writing systems have produced results that challenge the above assumptions. Specifically, these studies suggest that the characteristics of the orthographic system representing languages, and in particular orthographic consistency, i.e., the consistency of the mapping between graphemes and phonemes, impacts on the development of phonological recoding skills, which are basic to the acquisition of reading skills in all alphabetic orthographies (Share, 1995).

The role of orthographic transparency in first- and second-language reading acquisition

Alphabetic orthographies differ in the complexity of their grapheme-phoneme correspondences (*GPC*), or conversion rules. In *shallow* or *transparent* orthographies, the GPC are highly consistent, i.e., most of the graphemes always correspond to the same phonemes and vice versa; whereas in *deep, opaque* or *nontransparent* orthographies, these correspondences are rather inconsistent and unpredictable, i.e., many graphemes may be pronounced in several different ways, and/or many phonemes may be represented by several different graphemes.

If orthographic consistency impacts on the development of phonological recoding skills (cf. Share, 1995), then the development of the phonological decoding procedure using assembled pronunciations should be achieved earlier and more efficiently in transparent orthographic systems than in opaque ones. Indeed, in opaque orthographic systems, it is often necessary to rely on orthographic representations to supplement the processes of phonological assembly (Frith, Wimmer, & Landerl, 1998). The empirical research reviewed below supports this prediction.

Most of the cross-linguistic comparisons of reading acquisition in two languages differing in orthographic consistency have contrasted English with a more transparent orthographic system. For example, Wimmer and Goswami (1994) compared reading performances of seven- and nine-years-old English-speaking and (Austrian) German-speaking children. They observed a strong advantage in both groups of Austrian children compared to the English-speaking children in a task consisting in reading pseudowords, i.e., pronounceable but meaningless sequences of letters. However, the method of teaching reading differed somewhat across the two groups of children. Indeed, the school attended by the Austrian children used a rather systematic phonics approach, while the school attended by the English children used a

combination of phonics and whole-word reading schemes. These differences in instructional approaches may account for the observed pattern of results. Nevertheless, Landerl (2000) showed that although the performance of English-speaking first- and second-graders was better when these children were taught with phonics rather than with mixed methods of instruction, these children were still outperformed by German-speaking children. Such an advantage of German-speaking children over English-speaking children in reading pseudowords was also observed up until Grade 4 in a study conducted by Frith et al. (1998). The advantage persisted even when word-recognition abilities was equated between the two groups, and even when the children were presented with exactly the same items across languages. This was made possible because German and English share many words with similar spellings, pronunciations and meanings. Converging data have been reported in studies comparing English with Spanish and French (Goswami, Gombert, & de Barrera, 1998), as well as with Greek (Goswami, Porpodas, & Wheelwright, 1997). A study conducted by Seymour, Aro, and Erskine (2003) examined differences in the rates of acquisition of the components of foundation literacy in English and 12 other European languages which vary significantly in syllabic complexity and orthographic consistency. Their results are also consistent with the hypothesis that basic decoding skills develop slower and less effectively in deep than in shallow orthographies.

The results of simulations conducted by Hutzler, Ziegler, Perry, Wimmer, and Zorzi (2004) are also consistent with the empirical data reported above. Indeed, their cross-language implementation of a two-layer associative network (Plaut, McClelland, Seidenberg, & Patterson, 1996) was able to simulate the large initial advantage for pseudoword reading of the regular (German) over the irregular (English) orthographic system, but this was the case only when cross-language differences in teaching methods (phonics vs. whole-word approach) were taken into account by applying a teaching regime that specifically imitated the phonics approach typically found in regular orthographies. In accordance with the empirical data reported by Landerl (2000), the simulations suggest that while the English network does benefit from a phonics pre-training regime, this benefit is smaller and more restricted to early learning phases than with the German network.

Ellis and Hooper (2001) compared literacy acquisition in English and in Welsh, a language with transparent orthographic system. In accordance with the results reported above, they found that, in Grade 2, Welsh-speaking children read aloud better (61% correct tokens, 1821 types) than English-speaking children (52% correct tokens, 716 types), despite the fact that both groups of children were taught with very similar methods of reading instruction. They also made various observations suggesting that the Welsh-speaking readers relied more on an alphabetic decoding system than the English-

speaking readers. In particular, word length determined 70% of reading latencies in Welsh, against only 22% in English. The greater effect of word length in Welsh-speaking children suggests that these children assembled pronunciations by means of a left-to-right parse of the graphemes that constituted each word. Conversely, the fact that the English-speaking children were less affected by word length suggests that these children were using other cues to read aloud, probably orthographic. The different nature of the reading errors observed in the two groups is consistent with the hypothesis that the two groups of children used different reading strategies. Indeed, the Welsh-speaking children tended to produce pseudowords, whereas the English-speaking children tended to make real word substitutions and null attempts. The same pattern of results was reported by Spencer and Hanley (2003, 2004). Their study suggests, in addition, that learning to read in a transparent orthographic system increases phoneme awareness skills from the earliest stages of reading development on. Indeed, the Welsh-speaking readers performed better on phoneme awareness tasks than the English-speaking readers. Since the phonotactic structures of Welsh and English are similar, this difference is unlikely to be a consequence of differences in the syllabic structures of the two languages.

It is worth noting that differences in the development of reading procedures are also observed between orthographic systems which are less contrasted than the ones described in the studies reported above, for example between Spanish and Portuguese. Although both are considered as having transparent orthographic systems, grapheme-phoneme mappings are more consistent in Spanish than in Portuguese. As would be predicted on the basis of this difference, Spanish children were found to read pseudowords better than Portuguese children (Defior, Martos, & Cary, 2002).

Taken together, these findings suggest that the development of phonological recoding processes is slower and more difficult in less transparent orthographic systems than in more transparent orthographic systems. Moreover, these two types of systems seem to entail the use of different reading strategies. When GPCs are simple and straightforward, the development of phonological recoding is fast, and the mastery of phonemic assembly is usually sufficient for accurate word recognition. When GPCs are complex and irregular, the beginning reader has to supplement grapheme-phoneme conversion strategies with the use of larger units (e.g., rime) or with attempts at recognising whole words on the basis of partial cues (Goswami, Ziegler, Dalton, & Schneider, 2003).

The notion that phonological recoding skills develop with relatively more ease in transparent orthographic systems than in less transparent orthographic systems is also supported by studies conducted on reading acquisition in a L2, even in children who have very little linguistic proficiency in

the more transparent language. For example, Geva and Siegel (2000) examined the reading skills of 245 first- to five-graders taught to read and write concurrently in English, their L1, and in Hebrew, their L2. Unlike English, the vowelled Hebrew script can be considered as transparent in that there is a very consistent correspondence between graphemes and phonemes. The pronunciation of syllables in Hebrew varies rarely as a function of their position in the word. Therefore, unlike in English, the acquisition of GPC rules and decoding in vowelled Hebrew requires the learner to master few rules and few "exception words". Geva and Siegel (2000) found that children reached higher accuracy levels in decoding in Hebrew (L2), which has a more transparent writing system, than in English (L1). These differences were observed despite the obvious advantage of the children in L1 in terms of size of the lexicon and syntactic knowledge. Moreover, Geva and Siegel (2000) observed that the children's type of decoding errors were specific to the orthographic system. That is, younger children were more prone to make similar-word errors in English than in Hebrew. In Hebrew, younger children were able to decode many unfamiliar words with accuracy but without the appropriate stress. Word-stress is an essential element in Hebrew, and changes in stress may alter word meaning fundamentally. According to Geva and Siegel (2000), errors on the stress pattern of words reflect linear right-to-left syllable-based decoding, which have pronunciations and meanings that are presumably unfamiliar to the child. These findings suggest that there may be significant potential benefits to learn to read in the most transparent orthographic system first (vowelled Hebrew, as opposed to English), since the very consistent mappings from graphemes to phonemes in that language enhances phonological recoding skills, which are the essence of successful reading acquisition (Share, 1995; Share & Stanovich, 1995).

Transfer of reading procedures across languages

The idea that it might be beneficial to learn to read in the most consistent orthographic system, even if oral skills in that language are still weak, is further supported by the finding that beginning readers transfer word recognition skills from the more transparent orthographic system to the more opaque one. For example, a study of Carlisle and Beeman (2000) examined literacy acquisition in children of Hispanic background taught to read and write either in Spanish first (transparent orthographic system) or in English first (opaque orthographic system). In the 1st and 2nd Grades, these children were given standardised tests assessing oral language as well as reading and writing skills. They found that the children taught to read in Spanish first scored better in reading and writing tasks in Spanish, and did not differ in reading and writing tasks in English, compared to the children taught to read in English first. In

the same vein, Da Fontoura and Siegel (1995) found that bilingual Portuguese-Canadian reading-disabled children displayed significantly higher scores than monolingual English-speaking reading-disabled children in tasks consisting in reading and "spelling" pseudowords in English. According to Da Fontoura and Siegel (1995), these results might reflect a positive transfer of decoding skills across languages, which could result from the more regular GPC rules of Portuguese. However, in their study, this effect was not observed in normally achieving bilingual readers. Still, in a similar study conducted by D'Angiulli, Siegel, and Serra (2001) on Italian-English bilinguals, a positive transfer was reported both in skilled readers and in less skilled readers.

Mumtaz and Humphreys (2001) reported both positive and negative transfer from a transparent orthographic system to an opaque one. Indeed, they observed that Urdu-English bilingual children from Grade 2 and Grade 3 outperformed their English monolingual peers in reading tasks involving English regular word and pseudoword. Along with previous studies (Carlisle & Beeman, 2000; Da Fontoura & Siegel, 1995; D'Angiulli et al., 2001), these results suggest that beginning to read in a transparent orthographic system positively impacts on the development of phonological processes in the opaque orthographic system. However, Mumtaz and Humphreys (2001) observed a negative transfer, i.e., lower performances in the bilinguals than in the monolinguals, for English irregular words, which manifested itself through many errors consisting in regularisation. According to the authors, this was due to a greater reliance on non-lexical processing in the bilingual than in the monolingual children.

To summarise, research conducted in different linguistic contexts casts some doubts on the notion that the level of oral proficiency in a L2 might play a pervasive role in the development of basic word recognition skills in that language. Moreover, several studies highlight the importance of taking orthographic transparency into account when considering the concurrent development of reading and writing in two languages.

The present study examined the relative impact of oral proficiency and of orthographic transparency on reading development in French-native children immersed in Dutch and learning to read first either in that language (L2, but most transparent orthographic system) or in French (L1, but least transparent orthographic system). The present study also looked at the positive and/or negative transfer of word recognition strategies across languages in these two groups.

Characteristics of the Dutch and French orthographic systems

Because the study reported in the present contribution is concerned with the concurrent development of reading in French and Dutch, the section

below provides a brief overview of the orthographic systems of these two languages.

The Dutch orthographic system

Although the Dutch orthographic system is highly regular, several deviations from a one-to-one correspondence occur. Dutch orthography thus obeys a limited number of principles and rules that complement and restrict one another. These main principles, described below, are *economy*, *etymology*, and *uniformity* (Zonneveld, 1978).

Dutch has 16 vowels (Booij, 1995): five are short (/I/, /ɛ/, /ɔ/, /Y/, /ɑ/), seven are long (/i/, /y/, /u/, /e/, /ø/, /o/, /a/), one is the schwa (/ə/), and three are diphthongs (/ɛi/, /œy/, /ɔu/). There are only five letters (i, u, e, o, a) for the 13 Dutch (non diphthong) vowels. Whereas the spelling of short vowels is straightforward, the spelling of long vowels is more complicated. The generalisation (principle of economy) is that long vowels are spelled as single letters in open syllables and as geminates in closed syllables, that is, in syllables in which the vowel is followed by at least one consonant. Therefore, the word *raam* (window) is spelled *ramen* in its plural form. Moreover, a consonant between two vowels has to be duplicated if the preceding vowel is short and the second vowel is a schwa (e.g., the word *bot* [bɔt] (bone) is spelled *botten* in its plural form).

The geminate form of /i/ is *ie* rather than *ii*. In the case of /e/, it is also spelled as geminate *ee* in word-final position even though this is an open syllable, in order to avoid confusion with the schwa which is spelled as *e* in that position (Booij, 1995). Two other digraphs, *oe* and *eu*, represent the long vowel /u/ and /ø/. The three diphtongs of Dutch are spelled as sequences of two different letters (e.g., *ui* for /œy/), but complications arise because etymology plays a role in their spelling (Bos & Reitsma, 2003): /ɛi/ is spelled as *ei* when it derives historically from the Proto-Germanic /ɑi/, and as *ij* when it derives from the long /i/. Likewise, /ɔu/ is spelled as *ou* when derived from /ɔl/, but either as *ou* or *au* otherwise (Booij, 1995). Sometimes, the off-glide [ʊ] at the end of the diphthong is also represented in the spelling.

The spelling of the consonants is more straightforward. There are 18 letters to represent the 18 Dutch consonants (Booij, 1995). Still, spelling is complicated by the fact that voiced consonants become voiceless in final position. Thus *hard* (hard) and *hart* (heart) are both pronounced [hɑrt], and *lach* (smile, laughter) and *(ik) lag* (I laid) are both pronounced [lɑx]. In most of the cases, however, the spelling of these consonants obeys to the principle of uniformity: a root word is spelled in accordance with the spelling of its derivates (Bos & Reitsma, 2003; Zonneveld, 1978). For example, although the word *paard* (horse) is pronounced [pa:rt], it is spelled with a -d because its plural form is *paarden* (pronounced [pa:rd®n]). Similarly, *weg* is spelled

with a -g (although it is pronounced [wɛx]) because its plural form is *wegen* (pronounced [we:ɣən]). Thus, spellers can quite easily retrieve whether root words should be spelled with a letter corresponding to a voiced or voiceless segment on the basis of the derived forms.

Words from foreign origin which have become integrated into the vocabulary of Dutch (*loanwords*) were previously adapted to the rules of Dutch spelling or not, depending of the fact that their pronunciation would allow such an operation or not. Therefore, words such as *<u>k</u>ari<u>k</u>at<u>uur</u>* (caricature) but also *<u>c</u>onse<u>qu</u>ent* can be found, but other words, though, remained half-hearted, as in *ele<u>k</u>tri<u>c</u>ien* (Booij, 1995). Since 1998, only the written form from the language of origin is accepted in Belgium for loanwords.

The French orthographic system

French has 35 phonemes: 15 vowels, 17 consonants and three glides (Tranel, 1987). Like the Dutch alphabet, the French alphabet consists of 26 letters (five for vowels, 20 for consonants, and the semi-vowel y), but also has five diacritic marks (the cedilla, the acute accent, the grave accent, the circumflex accent and the dieresis), i.e., supplementary signs that, combined with some letters, form 13 additional symbols.

Although less transparent than the Dutch system, the French orthographic system is generally claimed to be relatively transparent for reading, with grapheme-phoneme associations quite predictable overall: more than 40% of these associations are completely regular and unambiguous (Lange & Content, 1999). However, some characteristics reduce its transparency when compared to the Dutch orthographic system. Indeed, there are more digraphs or complex graphemes that represent a single phoneme (93 graphemes, among which 57 of more than one letter, cf. Véronis, 1986) than in Dutch (35 graphemes, among which 13 of more than one letter, cf. Booij, 1995; Nunn, 1998). Moreover, compared to Dutch, there are more letters which may have different phonemic values, determined not only by general principles sensitive to context (e.g., the letter -c is pronounced [s] in front of the letters -e, -i, and -y, and [k] elsewhere), but also by idiosyncratic variations (e.g., -c has a third phonetic value, [g], that occurs in a few words: *zinc*, *second*). Additional complexity stems from final consonants (Tranel, 1987): if the presence of a (usually silent) final *e* is an infallible sign that the preceding consonant-letter must be pronounced, its absence does not mean that the final consonant is necessarily silent. For example, the letter *c* is usually pronounced like in *traffic* [trafik] (traffic) or *parc* [park] (park), but is silent in words like *estomac* [ɛstɔma] (stomach) or *porc* [pɔr] (pork). As Tranel (1987) also noted, the relationship between spelling and pronunciation in French is rendered still more complex by the fact that a letter or group of letters does not necessarily has a constant phonetic value within the same written

sequence. For instance, in the written sequence *ti* followed by a vowel, the letter *t* may be pronounced either [t], like in *sortie* [sɔrti] (exit) and *(nous) portions* [pɔrtjõ] (we were carrying), or [s], like in *inertie* [inɛrsi] (inertia) and *(les) portions* [pɔrsjõ] (the portions). As a result, any accurate rule-based description of the correspondences between graphemes and phonemes needs to incorporate a fairly large number of specific rules and exceptions (Yvon, de Mareuil, d'Allesandro, Auberge, Bagein, Bailly et al., 1998).

As far as spelling is concerned, French is considered to be more opaque than Dutch, given the complexity and ambiguity in the transcoding from sound to spelling. Ziegler and collaborators (Ziegler, Jacobs, & Stone, 1996; Ziegler, Stone, & Jacobs, 1997; see also Ziegler, Montant, & Jacobs, 1997) compared the degree of inconsistency both from spelling to phonology (*feedforward inconsistency*) and from phonology to spelling (*feedback inconsistency*). Their major result is that 79.1% of all monosyllabic French words are feedback inconsistent (their phonological body has more than one spelling) and 12.4% are feedforward inconsistent (their spelling body has more than one pronunciation). This lack of concordance between the spoken form and the spelling of many French words may readily be seen in a word like *vingt* (Galland, 1941). In spite of the fact that this word is composed of five letters, its pronunciation consists of only two distinct sounds: /v/ and /Ê/. Furthermore, if this spoken form is compared with its spelling, there is direct mapping only for *v*, as there is no /i/, /n/, /g/, or /t/ phoneme. In addition, many phonemes may be represented by different graphemes. For example, the nasal vowel /Ê/ may be spelled as *in, ain, un, im, ein, en, ym, aim, yn* (Véronis, 1986), and the nasal vowel /ã/ may be spelled as *am, amp, an, anc, and, ang, ans, ant, aon, emps, ens, ent* (Ziegler et al., 1996). Therefore, there are many words that, although pronounced the same, are spelled differently. Conversely, other words that are pronounced differently are spelled the same, for example *(les) fils* [fis] (the sons) and *(les) fils* [fil] (the threads). Hence, many orthographic forms, rules and exceptions have to be memorised by French spellers.

The present study

We hypothesised that the outlined differences between the Dutch and the French orthographic systems, in particular in terms of degree of orthographic transparency, would entail differences in the acquisition of reading in the two languages in French-native children in immersion in Dutch (henceforth, *Im*) and taught to read and spell either in Dutch first (henceforth, *ImD*), or in French first (henceforth, *ImF*).

More specifically, given the greater consistency of the Dutch orthographic system, we expected that the ImD would rely to a greater extent on the

phonological decoding procedure, using assembled pronunciations, than the ImF, who would need to rely to a larger extent on orthographic representations to supplement phonological assembly in order to read accurately.

Furthermore, we expected that even though their oral proficiency in their L2 was rudimentary, the ImD would reach high levels of accuracy in decoding in that language because its orthographic system is not very demanding for decoding.

Our third prediction concerned the transfer of reading procedures across languages. We expected that beginning to read in Dutch, the most transparent orthographic system, would positively affect phonological processes in reading in French, the most opaque orthographic system. If this were the case, the ImD should read better than the ImF in Dutch, and should read at least as well as the ImF in French, despite the fact that they were taught to read and spell in that latter language two years after the ImF.

To test these predictions, the Im children were compared to age-matched French-speaking monolinguals (*MonoF*) and Dutch-speaking monolinguals (*MonoD*).

In Grade 1 and at the beginning of Grade 2, all children were presented with various tasks assessing reading abilities (word and pseudoword reading, sentence comprehension), in the first language of reading instruction for the Im children. Thus, the ImF were assessed in French and compared to the MonoF, and the ImD were assessed in Dutch and compared to the MonoD.

By the end of Grade 2, the two Im groups started to learn to read in their second language of reading instruction. They were therefore presented with various tasks examining their reading abilities in both French and Dutch (word and pseudoword reading, as well as text comprehension). Their performances in these tasks in these two languages were compared to the ones of the French and Dutch monolinguals, respectively. Moreover, transfer from the most consistent orthographic system, Dutch, to the most opaque one, French, was examined by comparing the reading performances of the ImD to the ones of the MonoF and ImF for regular and irregular French words.

Method

Participants

Sixty-one French-speaking children attending a Dutch immersion program participated in a three-years longitudinal study, from Grade 1 to Grade 3. The 33 ImF were taught to read and spell in French first (in Grade 1), and then in Dutch (from Grade 2 or Grade 3 on); the 28 ImD were taught to read and spell in Dutch first (in Grade 1), and then in French (from Grade 2 or

Grade 3 on). The study also included 19 age-matched MonoF and 17 MonoD. In each of these four groups, the children came from two different schools, in an attempt not to confound the group factor with the school/classroom factor. The schools and children were matched as much as possible for socio-economic status on the basis of a questionnaire.

A questionnaire examining home habits in terms of language(s) use was presented to every child in Grade 1. It included questions regarding the language that was mostly used when the child was conversing with his parents, siblings, grandparents and friends. The results showed that most of the monolinguals had little or no exposure to the other language, and that the majority of the ImF and ImD spoke French with their family and friends outside the school. Only a few parents of the children from these two groups were able to speak Dutch fluently, but they never spoke Dutch with their child at home.

The four groups were further matched on their nonverbal cognitive abilities measured in Grade 1 with a standardised test (the Raven's Progressive Matrices – coloured version, Schutzenberg & Mavré, 1981; $F < 1$). Oral proficiency was also assessed in both languages in the Im children. In Grade 1, receptive vocabulary was examined with a task in which the children were required to select, among four pictures, the one that corresponded to a word said aloud in the target language, French (the Peabody Picture Vocabulary Test-Revised – Dunn, Thériault-Whalen, & Dunn, 1993) or Dutch (Dutch adaptation of the test, cf. Goetry, 2002). In Grade 2, the children were required to name as quickly and as accurately as possible 20 pictures (10 frequent, 10 rare) selected from the Snodgrass and Vanderwart's database (1980), in each language. They were also presented with a standardised test examining their oral comprehension of 92 sentences arranged in order of increasing difficulty (*ECOSSE*, cf. Lecocq, 1996), in which they were asked to select, among four pictures, the one that corresponded to a sentence said aloud in the target language.

Table 1 displays the mean scores observed in the two Im groups on the tasks assessing oral development in French and Dutch. These scores were compared to the ones of the MonoF and MonoN, respectively. Both in Grades 1 and 2, the ImF and ImD performed at the same level as the MonoF in all the tasks assessing oral development in French, despite the fact they were being instructed most of the time in Dutch. In Dutch, however, the three groups did not perform at the same level (receptive vocabulary: $F(2, 81) = 67.9$, $p < .001$; picture naming: $F(2, 76) = 177.6$, $p < .001$; oral sentence comprehension: $F(2, 75) = 24.2$, $p < .001$). The two Im groups performed more poorly than the MonoD on all the tasks assessing oral development in Dutch, $p < .001$ in all cases. This is not surprising given the fact that, although Dutch was taught at school, the children had few, if any, opportuni-

ties to practice that language at home. The ImF performed more poorly than the ImD in all the tasks in Dutch (receptive vocabulary and picture naming: $p < .001$; oral sentence comprehension: $p < .005$).

Table 1
Average scores (in %) in the tests measuring oral proficiency in both languages for each immersion group compared to French and Dutch monolinguals (standard deviations in brackets)

	French				Dutch		
	ImF	ImD	MonoF		ImF	ImD	MonoD
Receptive Vocabulary (Grade 1)	80	84	90		17	32	71
	[16.4]	[19]	[18.3]		[7.6]	[18.7]	[15.4]
Picture naming (Grade 2)	92	89	88		26	56	90
	[4.9]	[8.6]	[5.3]		[9.1]	[15.1]	[5.1]
Oral sentence comprehension (Grade 2)	89	87	85		59	67	81
	[4.2]	[8.8]	[5.7]		[10.2]	[10.3]	[5.7]

A questionnaire (Goetry, 2002) aimed at assessing the proportion of phonics and whole-word activities given to the children during class was presented to each teacher in Grade 1. Table 2 shows the relative percentages of items corresponding to phonics activities among the total number of assertions corresponding to such activities, and the relative percentages of items corresponding to 'whole-word' activities among the total number of assertions corresponding to such activities, separately for each group and in each school. As can be seen, although all the schools introduced reading by using a combination of phonics and whole-word reading approaches, the emphasis on individual grapheme-phoneme correspondences was stronger in the schools attended by the children taught to read in Dutch first (ImD and MonoD) than in the schools attended by the children instructed to read in French first (ImF and MonoF).

In order to control for the possible influence of these difference in reading methods across schools and groups on the patterns of observed results, a *phonics score* was calculated for each teacher by dividing the number of items corresponding to phonics exercises s/he selected by the total number of items chosen among all the assertions.

Materials and procedure

Children were tested individually in a quiet room at school. For each task, the experimenter explained the instructions and checked that the child had understood it with some practice items. A brief description of each task is provided below.

Table 2
Proportion of "whole-word" vs. "phonic" exercises (in %) in the eight schools
(Grade 1)

	"Whole-word"	"Phonics"
ImF		
School 1	57	63
School 2	62	37
Average ImF	60	50
MonoF		
School 1	52	5
School 2	52	56
Average MonoF	52	31
ImD		
School 1	52	79
School 2	48	74
Average ImD	50	77
MonoD		
School 1	38	63
School 2	48	74
Average MonoD	43	69

Reading performances in the first language of reading instruction

Word and pseudoword reading (Grades 1 and beginning of 2). In Grade 1, the French and Dutch single word reading tasks included 12 highly frequent real words (6 short and 6 long) and 12 pseudowords (6 short and 6 long) each, presented in order of increasing difficulty. Across languages, the stimuli were closely matched on segments length (number of syllables, letters, and phonemes) and word frequency, defined for French on the basis of *BRULEX* (Content, Mousty, & Radeau, 1990) and *NOVLEX* (Lambert & Chesnet, 2001), and for Dutch on the basis of *Streeflijst woordenschat voor zesjarigen* (Schaerlaekens, Kohnstamm, & Lejaegere, 1999) and *CELEX* (Burnage, 1990).

At the beginning of Grade 2, the reading tasks included 72 items varying in terms of lexicality (48 words vs. 24 pseudowords), word frequency (24 frequent vs. 24 unfamiliar), segments length (36 short vs. 36 long), and complexity (36 items with simple syllabic structures vs. 36 items with complex ones). The words and pseudowords were presented in a pseudo-random order and the children were told about the fact they would be asked to read both types of items in the same task. Again, the stimuli were closely matched across languages on segments length and word frequency, as well as on the nature of the initial consonant (to avoid major phonetic biases in voice key response time measurements, e.g., Kessler, Treiman, & Mullennix, 2002; Rastle & Davis, 2002).

In all of these reading tasks, the stimuli were presented one by one in the

middle of a computer screen. The children were asked to read them aloud as quickly and as accurately as possible. Their responses were written down by the first author and recorded on a Mini-Disc. Latencies were measured for each item. Stimulus presentation and timing, as well as data collection, were controlled using a vocal key connected to the *Psyscope* button box and *1.1. PPC software* (Cohen, MacWhinney, Flatt, & Provost, 1993), running on a *Macintosh Powerbook 180*. Only the latencies for correctly read items were considered in the statistical analyses.

Written sentence comprehension (beginning of Grade 2). In this test consisting in 20 sentences with a missing word, the children were asked to select the appropriate word out of five possibilities in order to complete each sentence. The vocabulary and syntactic structures increased in complexity throughout the test. The sentences were closely matched across languages on length, vocabulary complexity, and syntactic structures. The test was discontinued after five minutes.

Reading performances in both languages

Word and pseudoword reading (end of Grade 2 and Grade 3). The reading tasks presented at the end of Grade 2 and in Grade 3 included lexicality, word frequency and segments length as variables. Moreover, in each language, half of the list contained items with graphemes specific to the target language and the other half contained items with graphemes common to both languages.

Reading of regular and irregular words in French (Grade 3). The children were presented with 48 regular and 48 irregular words, varying in terms of segments length and frequency. The words were presented in a pseudo-random order. Each irregular word was matched to a regular word including the same graphemes (e.g., cl*ef* [kle] vs. br*ef* [brEf]). Moreover, the regular and irregular words were matched on word frequency, segments length and nature of the two initial consonants (see Kessler et al., 2002; Rastle & Davis, 2002).

Text comprehension (Grade 3). One text was designed for each language on the basis of books used at school. The children were required to read the text and then to answer to 10 written questions related to the text and formulated in the same language as the text. The texts were closely matched across the two languages on length (number of words), vocabulary complexity (the keywords were matched on frequency across languages), and syntactic structures. Although the syntactic correctness of the children's answers was measured, only reading comprehension scores will be presented here.

Results

Reading achievement in the first language of reading instruction

Grade 1: Word and pseudoword reading

Table 3 displays the mean percentages of words and pseudowords read correctly, separately for the two Im groups and for the two groups of monolinguals. Only analyses on accuracy and error types will be presented here. Indeed, the number of correct answers for long items was too small to allow analysing latencies of production (henceforth, *latencies*).

Table 3
Average reading scores (in %) in each group in Grade 1 (standard deviations in brackets)

| | Words | | | Pseudowords | |
	Short	Long		Short	Long
French reading					
ImF	56	59		35	23
	[34.8]	[29.7]		[33.6]	[26.2]
MonoF	51	71		27	14
	[27.8]	[21.6]		[27.3]	[21.8]
Dutch reading					
ImD	64	39		65	36
	[28.7]	[37.9]		[30.9]	[37.8]
MonoD	92	82		92	62
	[14.6]	[23.2]		[16.9]	[32.6]

Both participant ($F1$) and item ($F2$) repeated measures analyses of variance (*ANOVA*) were run on the accuracy scores, separately for children learning to read in French (MonoF and ImF) vs. in Dutch (MonoD and ImD). Each ANOVA included the factors of Lexicality (words vs. pseudowords), Length (short vs. long items) and Group (monolinguals vs. Im children).

The ANOVAs conducted on the results gathered from the children learning to read in French showed that the ImF and MonoF performed at similar levels, $F1$ and $F2 \leq 1$. There was an effect of Lexicality, because words were read better than pseudowords, $F1(1, 51) = 63.1$; $F2(1, 20) = 33.8$, both $p < .001$. The interaction between Lexicality and Group was unreliable, $F1(1, 51) = 2.6$, $p > .10$; $F2(1, 20) = 6.6$, $p < .05$, with the ImF slightly outperforming the MonoF for the pseudowords, $F2(1, 11) = 8.5$; $p < .05$, but not for the words, $F2 < 1$. Length did not affect performance, $F1$ and $F2 < 1$, and induced no interaction with Group, $F1 < 1$ and $F2(1, 20) = 3.1$, $p = .10$.

The analysis on the children learning to read in Dutch showed, unsurprisingly, that the ImD read more poorly than the MonoD, $F1(1, 42) = 11.4$, $p <$

.005; $F2(1, 20) = 87.6$, $p < .001$. Lexicality did not affect performance, $F1(1, 42) = 1.9$ and $F2(1, 20) = 2.3$, both $p > .10$, and did not interact with Group, $F1 < 1$; $F2(1, 20) = 1.8$, $p > .10$. The main effect of Length was significant, $F(1, 42) = 50.1$; $F2(1, 20) = 39.7$, both $p < .001$, because short items were read better than long items. There was no interaction between Length and Group, $F1$ and $F2 \approx 1$.

In order to examine possible differences in reading strategies between the children learning to read in French vs. in Dutch, repeated measures ANOVAs took the Language of reading instruction (French vs. Dutch) into account in addition to the factors of Group, Lexicality and Length. These analyses showed a significant main effect of Language, $F1(1, 97) = 22.9$; $F2(1, 20) = 66.1$, both $p < .001$, with the children learning to read in Dutch outperforming those learning to read in French (on the average, 63% vs. 42%, respectively). There were also significant main effects of Lexicality, $F1(1, 97) = 73.6$; $F2(1, 20) = 26.1$, both $p < .001$, and Length, $F1(1, 97) = 36.8$, $p < .001$; $F2(1, 20) = 9.4$, $p < .01$, both qualified by an interaction with Language of reading instruction. The Lexicality×Language×Group and Length×Language×Group interactions were not significant, $F1$ and $F2 < 1$ in both cases.

As illustrated in Figure 1, the Language×Lexicality interaction reflects the fact that Lexicality affected the children taught to read in French more strongly than those taught to read in Dutch, with average effect sizes of 35% vs. 6%, respectively, $F1(1, 97) = 37.6$; $F2(1, 20) = 22.5$, both $p < .001$. Coherently, the children learning to read in Dutch outperformed the ones learning to read in French for pseudowords, $F1(1, 100) = 36.4$; $F2(1, 11) = 105.2$, both $p < .001$, but not for words, $F1 \approx 1$; $F2(1, 11) = 2.3$, $p > .10$.
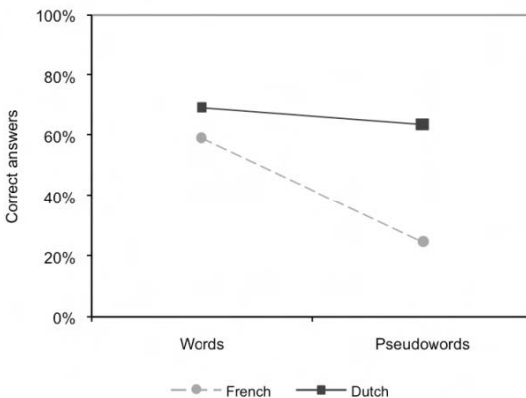

Figure 1

Average reading scores (in %) in the first language of reading instruction in Grade 1, separately for words and pseudowords and for French (gray dotted lines) and Dutch (black plain lines) readers in the first language of reading instruction

On the contrary, the Language×Length interaction reflects the fact that Length affected the children taught to read in Dutch but not those taught to read in French, with average effect sizes of 24% vs. 1%, respectively, $F1(1, 97) = 34.2$; $F2(1, 20) = 14.8$, both $p < .001$. As already illustrated above, and as shown in Figure 2, the children taught to read in French displayed equivalent performances for short and long items, whereas the children instructed to read in Dutch performed significantly better for the short than for the long items. Coherently, the latter outperformed the former for the short items, $F1(1, 100) = 34.6$; $F2(1, 11) = 63.1$, both $p < .001$, but not for the long items, $F1(1, 100) = 2.3$, $p > .10$; $F2(1, 11) = 3.3$, $p = .10$.
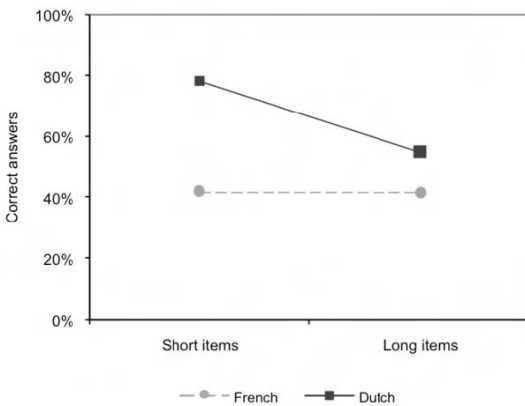
Figure 2

Average reading scores (in %) in the first language of reading instruction in Grade 1, separately for short and long items and for French (gray dotted lines) and Dutch (black plain lines) readers

Both the Language×Lexicality and the Language×Length interactions remained significant when methods of reading instruction were accounted for by using the phonics scores (see Method section) as a covariate in the analysis by participants, $F(1, 96) = 11.3$ and $= 16.5$, respectively, both $p \leq .001$. Thus, the differences observed between the children taught to read in French vs. in Dutch cannot be entirely accounted for by differences in the methods of reading instruction used by the teachers to teach reading in French vs. in Dutch.

Rather, the contrasted reading procedures adopted by the children taught to read in French vs. in Dutch may probably be related to the differing orthographic characteristics of French and Dutch. Indeed, because the children taught to read in Dutch (ImD and MonoD) displayed a small effect of lexicality but a strong effect of length, whereas the children taught in French (ImF and

MonoF) displayed the opposite pattern, the former groups seem to be more reliant on an alphabetic decoding strategy than the latter groups. The smaller effect of lexicality, together with the stronger effect of length, observed in the children taught to read in Dutch suggest that these children assembled pronunciations by means of a left-to-right parse of the graphemes that constituted each word. In contrast, the fact that the groups taught to read in French were not affected by word length, and showed poorer performances for pseudowords than the groups taught to read in Dutch, suggests that the former groups were less likely to attempt to construct pronunciations by the application of GPC rules than the latter groups, and tended to use other cues instead.

The qualitative analysis of the errors produced by the children taught to read in French vs. in Dutch provides further support for the hypothesis that the reading strategies differed across the two languages. Reading errors were classified into three categories: (i) null responses, (ii) whole-word substitutions, (iii) partial decoding or attempts that resulted in pseudowords. As can be seen in Figure 3, although in all children many errors consisted in decoding partially or producing pseudowords, the children taught to read in French produced relatively more whole-word substitutions than those taught to read in Dutch (on average, 37.4% of the total number of errors vs. 19.6% of this total, respectively); whereas the children taught to read in Dutch produced a majority of errors consisting in decoding partially or pseudowords (on average, 69.8%, compared to 47.6% in the children taught to read in French). This pattern of differences is significant, $\chi^2 = 22.5$, $p < .001$.
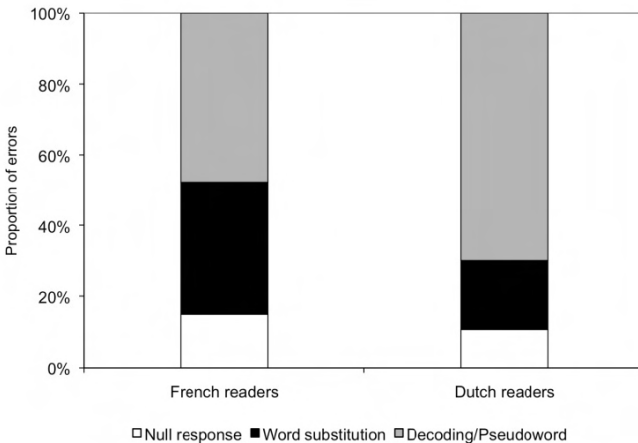

Figure 3
Proportion of null responses (in white), whole-word substitutions (in black) and nonword (decoding) responses (in gray), separately for French and Dutch readers reading in the first language of reading instruction in Grade 1

To summarise, in Grade 1, the children in immersion instructed to read in their native language first (ImF) read French at the same level as the French monolinguals, whereas, unsurprisingly, the children in immersion taught to read in their second language first (ImD) produced lower Dutch reading performances than the Dutch monolinguals, both for words and for pseudowords. However, the children instructed to read in Dutch (ImD and MonoD) read pseudowords better than the children taught to read in French (ImF and MonoF).

These two groups actually seemed to rely on different reading procedures adapted to the transparency of their instruction language: both the quantitative analyses on reading performances and the qualitative analyses of errors suggest that whereas the children taught to read in Dutch relied mostly on the phonological recoding procedure, those instructed to read in French seemed to supplement phonological assembly with other, probably lexical, cues.

Beginning of Grade 2: Word and pseudoword reading

Table 4 shows the percentages of correct responses for the various types of items in each group. Both participant and item ANOVAs were run on these data, separately for the children instructed in French vs. in Dutch. Each analysis included the factors of Lexicality/Frequency (frequent words vs. rare words vs. pseudowords), Length (short vs. long items), Complexity (simple vs. complex items) and Group (monolinguals vs. Im children).

Table 4
Average reading scores (in %) in each group in the beginning of Grade 2 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
French reading						
ImF						
Simple	96	82	79	54	78	41
	[7.0]	[20.3]	[19.1]	[25.4]	[20.2]	[25.7]
Complex	92	74	72	46	61	45
	[10.3]	[23.2]	[18.9]	[22.8]	[26.8]	[26.7]
MonoF						
Simple	98	81	84	68	70	44
	[7.6]	[24.3]	[17.9]	[23.8]	[20.5]	[34.7]
Complex	90	78	72	42	62	46
	[16.9]	[24.8]	[22.9]	[29.5]	[24.7]	[24.2]
Dutch reading						
ImD						
Simple	50	23	53	18	59	14
	[26.6]	[20.2]	[27.7]	[18.4]	[31.7]	[14.5]
Complex	75	27	70	23	72	27
	[30.8]	[25.4]	[30.6]	[21.8]	[33.2]	[30.7]
MonoD						
Simple	89	75	88	39	70	20
	[16.5]	[18.7]	[11.4]	[33.8]	[20.6]	[23.8]
Complex	96	65	88	26	88	40
	[7.4]	[22.6]	[12.8]	[20.4]	[14.0]	[17.8]

Like what had already been observed in Grade 1, the analysis on the children taught to read in French showed that the ImF and MonoF displayed similar levels of accuracy, $F1$ and $F2 < 1$. There was a significant main effect of Frequency, $F1(2, 96) = 72.9$; $F2(2, 60) = 43.2$, both $p < .001$, because frequent words were easier to read than rare words and pseudowords. Length also affected performances, $F1(1, 48) = 213.4$; $F2(1, 60) = 58.4$, both $p < .001$, since short items were easier to read than long ones. There was also a significant effect of Complexity, $F1(1, 48) = 32.3$, $p < .001$; $F2(1, 60) = 8.1$, $p < .01$, because simple items were easier to read than complex ones. None of these effects interacted with Group, all $p > .10$. Similar effects were also observed on the latencies (see Table 5), with no significant difference between the two groups, $F1 < 1$ and $F2(1, 60) = 3.3$, $p < .10$, but main effects of Frequency, $F1(2, 68) = 31.7$; $F2(2, 60) = 11.8$, and Length, $F1(1, 34) = 90.3$; $F2(1, 60) = 109.4$, all $p < .001$. The effect of Complexity, though, was not significant, $F1(1, 34) = 1.5$, $p > .10$; $F2 < 1$.

Table 5

Mean reading speed (in ms) in each group in the beginning of Grade 2
(standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
French reading						
ImF						
Simple	1402	2152	1683	2593	2024	3175
	[425]	[844]	[591]	[1171]	[912]	[1713]
Complex	1387	2275	1739	2831	1546	3540
	[423]	[1025]	[623]	[1491]	[620]	[2315]
MonoF						
Simple	1492	2438	1926	2669	2074	3811
	[466]	[741]	[559]	[1008]	[779]	[1462]
Complex	1330	2452	1755	2930	1615	3175
	[340]	[1205]	[603]	[1218]	[613]	[1414]
Dutch reading						
ImD						
Simple	1324	3377	1421	3181	1393	3511
	[434]	[1087]	[447]	[1030]	[372]	[1389]
Complex	1327	3306	1378	3844	1416	3605
	[328]	[1459]	[319]	[1883]	[351]	[1597]
MonoD						
Simple	938	1200	1467	1975	1174	2278
	[176]	[343]	[708]	[1227]	[281]	[1182]
Complex	1068	1434	1282	2351	1203	2093
	[240]	[604]	[469]	[1768]	[419]	[867]

In contrast, the data from the children instructed in Dutch showed that the ImD read more poorly, $F1(1, 40) = 12.9$; $F2(1, 60) = 135.4$, both $p < .001$, and more slowly, $F1(1, 11) = 6.2$, $p < .05$; $F2(1, 51) = 95.4$, $p < .001$, than the MonoD. There was a significant effect of Frequency, with frequent words being read

more accurately, $F1(2, 80) = 27.9$, $p < .001$; $F2(2, 60) = 6.6$, $p < .005$, and slightly faster, $F1(2, 22) = 4.5$, $p < .05$; $F2(2, 51) = 2.7$, $p = .07$, than rare words and pseudowords. Length also affected performance, with short items being read more accurately, $F1(1, 40) = 409.5$; $F2(1, 60) = 158.8$, both $p < .001$, and faster, $F1(1, 11) = 37.5$; $F2(1, 51) = 124.7$, both $p < .001$, than long items. The effect of Complexity was also significant in the analyses on accuracy, $F1(1, 40) = 33.6$, $p < .001$; $F2(1, 60) = 6.3$, $p < .05$, but in this case the complex items were easier to read than the simple ones. This may be a consequence of the Dutch phonotactic structure: complex syllabic structures are more frequent in this language than simple syllabic structures, and hence may favour reading.

In the analyses on accuracy, these main effects were qualified by a series of interactions with Group. Among these, the most interesting theoretically are the one between Group and Frequency, $F1(2, 80) = 21.3$; $F2(2, 60) = 15.2$, both $p < .001$ and the one between Group, Frequency and Length, $F1(2, 80) = 14.3$, $p < .001$; $F2(2, 60) = 4.7$, $p = .01$.

The Group×Frequency interaction, illustrated in Figure 4, reflects the fact that the ImD did not show any effect of Frequency in reading, $F1$ and $F2 \leq 1$, unlike the MonoD, $F1(2, 32) = 31.7$, $p < .001$; $F2(2, 69) = 6.1$, $p < .005$. Analyses conducted separately on the three types of items indicated that the ImD were clearly less accurate than the MonoD at reading both the frequent and the rare words, $F1(1, 43) = 39.6$, $p < .001$ and $= 9.7$, $p < .005$, respectively; $F2(1, 23) = 82.4$ and $= 21.5$, respectively, both $p < .001$, but only slightly less accurate than the MonoD at reading pseudowords, $F1(1, 43) = 3.3$, $p < .10$; $F2(1, 23) = 19.5$, $p < .001$.
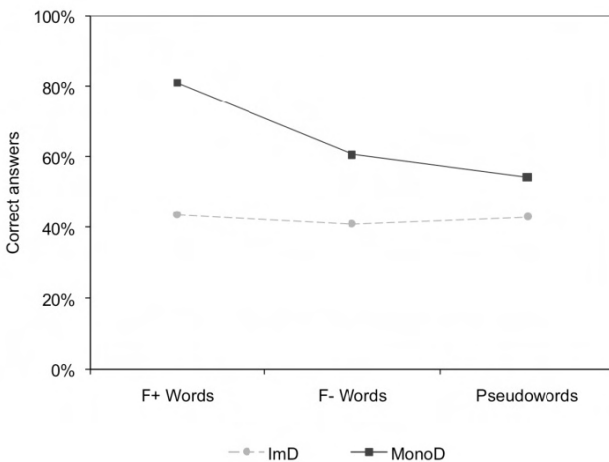

Figure 4

Average reading scores (in %) in the first language of reading instruction in the beginning of Grade 2, separately for frequent words, rare words and pseudowords reading and for the ImD (gray dotted line) and MonoD (black lain line) groups

The Group×Frequency×Length interaction, illustrated in Figure 5, was due to the fact that in the MonoD, the effect of length was much weaker for the frequent words than for the rare words and pseudowords, whereas in the ImD this effect was equivalent across the three types of items. This suggests that the MonoD were starting to use the lexical procedure to read familiar words, whereas the ImD still relied mostly on phonological recoding even for reading frequent words. This is probably related to the ImD's restricted levels of lexical knowledge in Dutch, as suggested by the strong correlations between their reading proficiency and vocabulary level in that language (see Table 6).
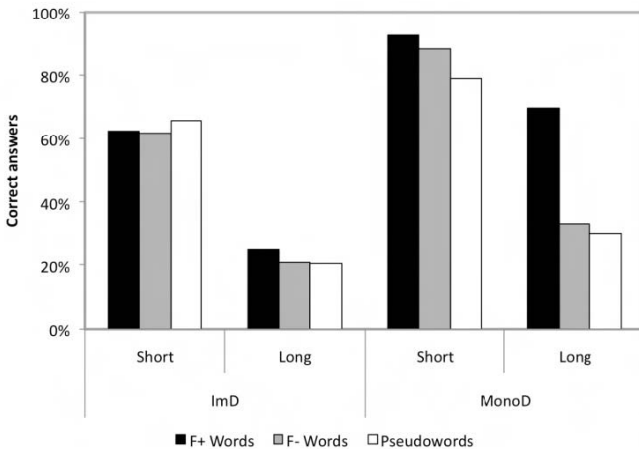
Figure 5
Average reading scores (in %) in the first language of reading instruction in the beginning of Grade 2, separately for frequent words (in black), rare words (in gray) and pseudowords (in white), for short and long items and for the ImD and MonoD groups

Table 6
Correlations in the ImD group between reading proficiency in Dutch and vocabulary level assessed in Grade 2 with a picture naming test and an oral sentences comprehension test

	Oral sentences comprehension	Picture naming
Total	.49*	.72**
Frequent words	.56*	.73**
Rare words	.46*	.69**
Pseudowords	.43*	.68**
* p < .05		
** p < .001		

To summarise, at the beginning of Grade 2, the Im children instructed to read in Dutch first had difficulties in exploiting lexical strategies in reading, probably because they still had restricted levels of linguistic proficiency in that language. Nevertheless, these children did not seem to experience greater difficulties than the MonoD in exploiting the GPCs of their *second* language to decode pseudowords. When comparing results from French and Dutch readers (see Tables 4 & 5), however, it appears that the MonoD group read long pseudowords less correctly than the MonoF (and ImF), at least for simple items. Given the fact that Dutch orthography is more transparent than the French one, the reverse result would have been expected. Still, it should be noted that the MonoD children read these items much more rapidly than did the MonoF and ImF. In addition, direct comparison of these performances is difficult to interpret since, contrary to the material used in first Grade, the pseudowords presented to French and Dutch readers were here different.

Beginning of Grade 2: Written sentences comprehension

The scores for written sentences comprehension were estimated with the following formula: {(total number of sentences correctly read in 5 minutes)/total number of test sentences) x100}. Table 7 shows the average scores, separately for the groups instructed to read in French vs. in Dutch. No significant difference was observed between the two groups of children taught in French, $t < 1$. In contrast, the ImD performed more poorly than the MonoD, $t(43) = 4.9$, $p < .001$.

Table 7
Average scores (in %) of the French and Dutch readers on the written sentence comprehension task presented in the beginning of Grade 2 (standard deviations in brackets)

	Written sentences comprehension
French	
ImF	66
	[22.7]
MonoF	68
	[22.0]
Dutch	
ImD	35
	[22.4]
MonoD	70
	[23.4]

Reading achievement in both languages

The two groups of Im children were tested in both French and Dutch by the end of Grade 2 and again in Grade 3. We first compared their levels of reading achievement to the ones of the French and Dutch monolinguals,

respectively. Next, we compared their reading performances in their two languages. However, because not all of the immersion schools enrolled in the present study started reading instruction in the other language by the end of Grade 2, the data collected at the end of Grade 2 only concern 16 ImD and 15 ImF. In Grade 3, all of the Im children were assessed in their second language of reading instruction.

Reading performance of the two groups in immersion in Dutch
End of Grade 2: Word and pseudoword reading in Dutch. Tables 8 and 9 display the average accuracy scores and latencies observed for the reading of words and pseudowords in Dutch. Both analyses by participants and by items were conducted on these data, including the factors of Frequency/Lexicality (frequent words vs. rare words vs. pseudowords), Length (short vs. long items), Specificity (graphemes common to both languages vs. specific to Dutch) and Group (ImD, ImF, MonoD).

The analysis conducted on accuracy scores showed that the three groups did not perform at the same level, $F1(2, 43) = 9.8$; $F2(2, 120) = 54.8$, both $p < .001$. Post-hoc tests showed that the ImF displayed lower performances than both the ImD and MonoD, $p < .01$ and $< .005$, respectively; the latter two groups did not differ significantly from each other, $p > .10$. However, care must be taken in the interpretation of the difference between the performance of the ImF and the ImD, since the analyses on latencies showed a significant, although unreliable, effect of Group, $F1(2, 26) = 2.6$, $p < .10$; $F2(2, 118) = 41.9$, $p < .001$, indicating that the ImD actually read somewhat more slowly than both the MonoD and ImF, both $p < .001$ (the latter two groups did not differ from each other, $p < .10$).

Table 8
Average reading scores in Dutch (in %) for the ImF, ImD and MonoD groups in the end of Grade 2 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImF						
Common	66	41	79	58	79	54
	[19.4]	[33.8]	[19.4]	[27.4]	[24.7]	[31.1]
Specific	61	33	51	37	58	22
	[28.6]	[19.9]	[22.3]	[21.1]	[31.2]	[21.4]
ImD						
Common	81	52	93	68	84	56
	[21.8]	[25.1]	[10.5]	[27.4]	[19.7]	[33.1]
Specific	84	74	94	61	75	36
	[14.2]	[23.4]	[10.3]	[33.6]	[16.0]	[25.9]
MonoD						
Common	91	63	81	70	78	51
	[12.4]	[26.8]	[15.1]	[19.1]	[19.5]	[27.0]
Specific	99	83	90	74	82	48
	[4.4]	[15.3]	[12.3]	[28.6]	[11.6]	[17.8]

Table 9
Mean reading speed in Dutch (in ms) for the ImF, ImD and MonoD groups in the end of Grade 2 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImF						
Common	1261	1782	1353	1864	1220	1517
	[645]	[696]	[445]	[1187]	[401]	[568]
Specific	1179	1407	1344	1396	1200	1339
	[379]	[749]	[461]	[489]	[443]	[811]
ImD						
Common	1521	2977	1797	3327	1680	3219
	[843]	[1527]	[744]	[1778]	[623]	[1699]
Specific	1335	2334	1741	2826	1808	3285
	[385]	[1043]	[798]	[1391]	[953]	[1833]
MonoD						
Common	1297	2006	1401	2380	1803	3004
	[916]	[1282]	[717]	[1268]	[977]	[1802]
Specific	976	1505	1033	2131	1275	2273
	[189]	[709]	[294]	[828]	[471]	[1322]

Frequency affected performances, with frequent words being read more accurately, $F1(2, 86) = 25$, $p < .001$; $F2(2, 60) = 4.8$, $p = .01$, and somewhat more rapidly, $F1(2, 52) = 12.8$, $p < .001$; $F2(2, 59) = 2.3$ $p < .10$, than rare words and pseudowords. Length also modulated performances, because short items were read more accurately, $F1(1, 43) = 156.1$; $F2(1, 60) = 64.3$, both $p < .001$, and more rapidly, $F1(1, 26) = 47.8$; $F2(1, 59) = 97.1$, both $p < .001$, than long ones. Specificity led to a speed-accuracy trade-off, with graphemes common to both languages read better, $F1(1, 43) = 7.7$, $p < .001$; $F2(1, 60) = 2.3$, $p > .10$, but more slowly, $F1(1, 26) = 12.5$; $F2(1, 59) = 10.2$, both $p < .005$, than graphemes specific to Dutch.

The interaction between Group and Frequency was significant both in the analyses on accuracy, $F1(4, 86) = 9$, $p < .001$; $F2(4, 120) = 4.5$, $p < .005$, and latencies, $F1(4, 52) = 3.2$; $p < .05$; $F2(4, 118) = 5.1$, $p = .001$. The Frequency effect reached significance in the ImD (accuracy: $F1(2, 30) = 25.5$, $p < .001$; $F2(2, 69) = 3.0$, $p = .055$; latencies: $F1(2, 22) = 7.9$, $p < .005$; $F2 < 1$) and MonoD (accuracy: $F1(2, 28) = 22.3$, $p < .001$; $F2(2, 69) = 6.8$, $p < .005$; latencies: $F1(2, 18) = 12.1$, $p < .001$; $F2(2, 68) = 5.9$, $p < .005$). Closer inspection of Tables 8 and 9 showed that, contrary to the MonoD, who obtained a regular word frequency effect, the ImD tended to be more accurate on rare than on frequent words, $F1(1, 15) = 8.9$, $p < .01$; $F2 < 1$. However, they read frequent words faster than rare words, $F1(1, 11) = 9.3$, $p = .01$; $F2 < 1$, and, like the MonoD, they read better both frequent and rare words than pseudowords, $F1(1, 15) = 26.7$, $p < .001$; $F2(1, 47) = 1.9$, $p > .10$ and $= 33.7$, $p < .001$; $F2(1, 47) = 6.2$, $p < .05$, respectively. This was not the case of the ImF, who obtained equivalent performances and latencies for either frequent, rare words or pseudowords (Frequency effect not significant on accuracy:

$F1(2, 28) = 1.6$; $F2 < 1$; and latencies: $F1 < 1$; $F2(2, 68) = 1.8$, $p > .10$).

Analyses conducted separately for each type of items showed that the ImF displayed lower scores than the two other groups for both frequent words ($F1(2, 45) = 17.4$; $F2(2, 46) = 40.3$, both $p < .001$) and rare words ($F1(2, 45) = 10$; $F2(2, 46) = 14.5$, both $p < .001$). For pseudowords, the effect of Group was significant only in the item analysis on accuracy scores, $F1(2, 45) = 2.1$, $p > .10$; $F2(2, 46) = 4.4$; $p < .05$. Similar but unreliable effects of Group were observed on latencies.

The Group×Specificity interaction was significant in the accuracy analyses, $F1(2, 43) = 22.1$; $F2(2, 120) = 16.6$, both $p < .001$. The ImF group read more poorly items with graphemes specific to Dutch than items with graphemes common to both languages, $F1(1, 14) = 26.9$; $F2(1, 70) = 12.9$, both $p = .001$, whereas the ImD group displayed similar performances for both materials, $F1$ and $F2 < 1$. The MonoD group tended to read better the items with graphemes specific to Dutch than those with graphemes common to both languages, $F1(2, 28) = 22.3$, $p < .001$; $F2(1, 70) = 2.1$, $p > .10$.

To summarise, by the end of Grade 2 the ImF performed more poorly than both the ImD and MonoD in reading words and pseudowords in Dutch. In particular, items containing graphemes specific to Dutch represented a major difficulty for them. On the contrary, the ImD read as accurately as the MonoD, although more slowly than the latter. However, with respect to the effect of word frequency, neither the ImD nor the ImF children behaved like the Dutch monolinguals. Indeed, unlike the MonoD, neither of the two Im groups benefited from the orthographic and phonological familiarity of frequent words. Still, similarly to the MonoD, the ImD read better both frequent and rare words than pseudowords. This was not the case of the ImF, who read words at the same level as pseudowords.

Grade 3: Word and pseudoword reading in Dutch. Tables 10 and 11 display the average accuracy scores and latencies observed in the tasks involving the reading of words and pseudowords in Dutch.

Analyses including Frequency/Lexicality (frequent words vs. rare words vs. pseudowords), Length, Specificity and Group as factors, indicated that the three groups did not perform at the same level of accuracy, $F1(2, 57) = 23.9$; $F2(2, 120) = 98.7$, both $p < .001$ and speed, $F1(2, 37) = 4.3$, $p < .05$; $F2(2, 118) = 126.6$, $p < .001$. Like at the end of Grade 2, post-hoc tests indicated that the ImF displayed lower performances compared to both the MonoD and ImD (accuracy: both $p < 001$; speed: both $p < .05$), whereas the latter two groups reached similar levels of performance, all $p > .10$.

Table 10

Average reading scores in Dutch (in %) for the ImF, ImD and MonoD groups in Grade 3 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImF						
Common	74	43	79	64	73	63
	[20.9]	[19.6]	[11.9]	[20.2]	[20.4]	[24.6]
Specific	70	45	65	38	59	34
	[32.7]	[22.5]	[22.4]	[27.2]	[32.7]	[22.4]
ImD						
Common	95	61	93	78	91	74
	[9.2]	[20.9]	[9.9]	[17.3]	[14.3]	[22.3]
Specific	90	79	98	76	85	57
	[11.1]	[17.8]	[5.0]	[15.9]	[11.3]	[21.6]
MonoD						
Common	98	87	93	85	93	76
	[5.7]	[11.1]	[12.1]	[12.9]	[12.1]	[22.2]
Specific	98	87	96	83	93	63
	[5.7]	[13.8]	[7.5]	[14.3]	[8.9]	[18.2]

Table 11

Mean reading speed in Dutch (in ms) for the ImF, ImD and MonoD groups in Grade 3 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImF						
Common	1433	2132	1446	1970	1447	2087
	[414]	[1040]	[506]	[984]	[553]	[1010]
Specific	1510	1933	1677	2702	1379	2128
	[600]	[1108]	[582]	[1835]	[486]	[1308]
ImD						
Common	1100	2050	1201	1962	1285	2201
	[341]	[1329]	[485]	[750]	[567]	[983]
Specific	990	1358	1856	1948	1172	2218
	[224]	[499]	[273]	[903]	[324]	[1218]
MonoD						
Common	703	1343	790	1062	1116	1498
	[124]	[843]	[153]	[256]	[280]	[337]
Specific	657	751	759	958	1094	1358
	[45]	[82]	[120]	[152]	[423]	[387]

As in Grade 2, the main effect of Length was significant in the analyses on both accuracy ($F1(1, 57) = 197.9$; $F2(1, 60) = 40.9$, both $p < .001$) and latencies ($F1(1, 37) = 34.2$; $F2(1, 59) = 100.3$, both $p < .001$). The main effect of Frequency was unreliable both on accuracy, $F1(2, 114) = 12.9$, $p < .001$; $F2 (2, 60) = 2.2$, $p > .10$, and latencies, $F1(2, 74) = 8.6$, $p < .001$; $F2(2, 59) = 2.6$, $p = .08$.

Specificity affected accuracy, $F1(1, 57) = 17.2$, $p < .001$; $F2(1, 60) = 3.6$, $p = .06$, and interacted with Group in both the analyses on accuracy, $F1(2, 57) = 6.4$; $F2(2, 120) = 5.6$, both $p \le .005$, and latencies, $F1(2, 37) = 5.1$, $p < .05$;

$F2(2, 118) = 8.8$, $p < .001$. The ImF read words with graphemes common to both languages more accurately, $F1(1, 30) = 19.8$, $p < .001$; $F2(1, 70) = 7.5$, $p < .01$, and slightly faster, $F2(1, 69) = 3.6$, $p = .06$; but $F1(1, 14) = 1.9$, $p > .10$, than words with graphemes specific to Dutch, which was not the case of the two other groups, all $F1$ and $F2 < 1$; on the contrary, the latter tended to read words with graphemes specific to Dutch faster than words with graphemes common to both language (MonoD: $F1 (1, 6) = 4.4$, $F2 (1, 69) = 2.8$, both $p < .10$; ImD: $F1(1, 17) = 10.3$, $p < .01$; $F2 < 1$).

The interaction between Group and Frequency was significant in the analysis on latencies, $F1(4, 74) = 4.3$, $p < .005$; $F2(4, 118) = 9.4$, $p < .001$, and approached significance in the analysis on accuracy by participants, $F1(4, 114) = 2.3$, $p = .06$; $F2(2, 120) = 1.4$; $p > .10$. The effect of Frequency was not significant in the ImF group (accuracy: $F1(2, 60) = 1.9$, $p > .10$; latencies: $F1(2, 28) = 2.7$, $p < .10$; $F2(2, 68) = 1.3$, $p > .10$), unlike in the other two groups (accuracy: $F1(2, 42) = 12.7$, $p < .001$ for the ImD; $F1(2, 16) = 7.1$, $p < .01$ for the MonoD; latencies: $F1(2, 34) = 7.9$, $p < .005$; $F2(2, 68) = 3.05$, $p = .05$ for the ImD; $F1(2, 12) = 10.7$, $p < .005$; $F2(2, 68) = 11.9$, $p < .001$ for the MonoD). As in Grade 2, the ImD read better rare words than frequent words, $F1(1, 21) = 11.6$, $p < .005$, but read frequent words faster than rare words, $F1(1, 17) = 5.5$, $p < .05$; $F2(1, 46) = 1.6$; $p > .10$, and, like the MonoD, read better both frequent and rare words than pseudowords, $F1(1, 21) = 4.8$, $p < .05$ and $= 22.5$, $p < .001$, respectively.

Analyses conducted on accuracy separately for each type of items indicated a significant Group effect in all cases, $F1(2, 61) = 31.6$, $= 36.2$, and $= 12.1$, for frequent words, rare words and pseudowords, respectively, all $p < .001$. Indeed, the ImF performed more poorly than the other two groups on frequent and rare words, all $p < .001$, as well as on pseudowords, all $p < .005$. Almost similar effects were observed on latencies, with the effect of Group being significant for frequent words, $F1(2, 39) = 5.8$, $p < .01$; $F2(2, 44) = 37.3$; $p < .001$, and for rare words, $F1(2, 39) = 6.4$; $p < .005$; $F2(2, 46) = 59.9$, $p < .001$, but unreliable for pseudowords, $F1(2, 39) = 1.4$; $p > .10$; $F2(2, 46) = 12.8$; $p < .001$. Indeed, the ImF read both the frequent and the rare words slower than the MonoD, $p < .01$ and $< .005$, respectively, and the two Im groups read the pseudowords slower than the MonoD, both $p < .001$.

To summarise, in Grade 3, the ImF still performed more poorly than both the ImD and MonoD. As was already the case at the end of Grade 2, the ImF experienced greater difficulties with graphemes specific to Dutch than with graphemes common to both languages. In contrast, the ImD did not differ from the MonoD in terms of level of proficiency and latencies. However, as in Grade 2, the two Im groups did not behave like the Dutch monolinguals with respect to word frequency. Indeed, unlike the MonoD, neither of them obtained higher scores on frequent words than on rare ones.

Grade 3: Text comprehension in Dutch. Table 12 shows the mean percentages of correct answers in the text comprehension task in Dutch presented in Grade 3. The three groups did not perform at the same level, $F(2, 57) = 5.7$, $p < .01$. Post-hoc analysis indicated that the ImF displayed significantly lower scores than both the MonoD and ImD, both $p < .05$, while the ImD did not differ from the MonoD, $p > .10$.

Table 12

Average comprehension scores (in %) in the Dutch text comprehension task presented in Grade 3, separately for the ImF, ImD and MonoD groups (standard deviations in brackets)

	Text comprehension
ImF	76
	[14.1]
ImD	87
	[12.6]
MonoD	91
	[7.8]

Reading performance of the two immersion groups in French

End of Grade 2: Word and pseudoword reading in French. Tables 13 and 14 present the mean accuracy scores and latencies for reading in French, separately for the three types of items (frequent words, rare words, pseudowords) and the three groups (ImD, ImF, MonoF). The analyses included the same factors as for the reading tasks in Dutch.

Table 13

Average reading scores in French (in %) for the ImD, ImF and MonoF groups in the end of Grade 2 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImD						
Common	88	75	82	63	77	47
	[16.6]	[30.9]	[22.2]	[33.6]	[24.3]	[33.9]
Specific	78	57	64	64	69	44
	[29.5]	[35.9]	[26.7]	[37.5]	[35.4]	[32.7]
ImF						
Common	96	92	76	64	77	64
	[7.7]	[12.3]	[22.6]	[24.2]	[21.6]	[25.0]
Specific	97	79	76	89	80	63
	[7.0]	[17.1]	[18.7]	[8.3]	[21.1]	[18.6]
MonoF						
Common	96	86	90	66	75	63
	[8.9]	[14.9]	[16.9]	[22.5]	[21.0]	[23.9]
Specific	93	76	73	82	85	55
	[13.9]	[19.4]	[19.4]	[19.6]	[14.5]	[26.0]

Table 14
*Mean reading speed in French (in ms) for the ImD, ImF and MonoF groups
in the end of Grade 2 (standard deviations in brackets)*

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImD						
Common	1576	2374	1695	3020	2022	3004
	[732]	[1293]	[763]	[1718]	[815]	[1421]
Specific	1464	2566	1763	2897	1718	3547
	[681]	[1696]	[945]	[1680]	[704]	[1791]
ImF						
Common	1419	2517	1715	2761	1742	2559
	[883]	[2208]	[982]	[2463]	[966]	[2217]
Specific	1377	2343	1626	2332	1666	2617
	[745]	[1727]	[1094]	[2039]	[796]	[1850]
MonoF						
Common	1475	2542	1819	2690	1878	2373
	[538]	[891]	[578]	[1137]	[822]	[1063]
Specific	1356	2111	1642	2213	1668	2816
	[431]	[751]	[518]	[792]	[473]	[1161]

There were significant main effects of Frequency, with frequent words read more accurately, $F1(2, 94) = 59.1$; $F2(2, 60) = 14.1$, both $p < .001$, and faster, $F1(2, 76) = 22.45$, $p < .001$; $F2(2, 60) = 6.6$, $p < .005$, than rare words and pseudowords, as well as of Length, with short items read more correctly, $F1(1, 47) = 57.9$; $F2(1, 60) = 23.1$, both $p < .001$, and more rapidly, $F1(1, 38) = 46.9$; $F2(1, 60) = 128.8$ both $p < .001$, than long items. The effect of Specificity was unreliable on accuracy, $F1(1, 47) = 5.7$, $p < .05$; $F2(1, 60) = 1.2$, $p > .10$, and did not affect latencies, $F1(1, 38) = 1.9$, $p > .10$; $F2 < 1$.

The effect of Group was also unreliable, both on accuracy, $F1(2, 47) = 2.46$, $p = .10$; $F2(2, 120) = 40.7$, $p < .001$, and latencies, $F1 < 1$; $F2(2, 120) = 8.2$, $p < .001$. However, the interaction between Group and Specificity was reliable in the analyses on accuracy, $F1(2, 47) = 6.3$, $p < .005$; $F2(2, 120) = 7.8$, $p = .001$ (latencies: $F1$ and $F2 \leq 1$). It was due to the Specificity effect being stronger for the ImD group than for the two other groups. Indeed, the ImD performed more poorly on items with graphemes specific to French than on items with graphemes common to both languages, $F1(1, 15) = 11.3$, $p < .005$; $F2(1, 70) = 5.4$, $p < .05$, which was not the case in the MonoF, $F1(1, 18) = 1.9$, $p > .10$; $F2 < 1$, and ImF, $F1$ and $F2 < 1$. Analyses conducted on accuracy separately for each type of items indicated a significant Group effect for those with graphemes specific to French, $F1(2, 49) = 4.3$, $p < .05$; $F2(2, 70) = 44.1$, $p < .001$, with the ImD performing more poorly than the other two groups, both $p \leq .05$. The Group effect was unreliable for the items with graphemes common to both languages, $F1 < 1$; $F2(2, 70) = 5.8$, $p = .005$.

To summarise, by the end of Grade 2, the overall reading level of the ImD in French, their native language, but second language of reading instruction,

was close to the one of the MonoF and ImF. Indeed, the group differences were reliable only for words including graphemes specific to French, for which the ImD encountered some difficulties.

Grade 3: Word and pseudoword reading in French. Table 15 and Table 16 show the mean accuracy scores and reading latencies, separately for the three types of items and for the three groups. No significant effect of Group was observed on accuracy, $F1 < 1$; $F2(2, 120) = 2.1$, $p > .10$. Still, there was an unreliable effect of this factor on latencies, $F1 < 1$; $F2(2, 120) = 12.1$, $p < .001$, with the ImD reading more slowly than both the MonoF and ImF, $p < .001$ and $p < .05$ respectively. The ImF also read more slowly than the MonoF, $p = .001$.

Table 15

Average reading scores in French (in %) for the ImD, ImF and MonoF groups in Grade 3 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImD						
Common	93	86	70	72	93	67
	[9.9]	[14.4]	[21.2]	[16.9]	[12.1]	[22.4]
Specific	74	91	90	76	84	69
	[17.2]	[11.1]	[13.1]	[19.9]	[16.9]	[21.9]
ImF						
Common	92	84	65	78	90	79
	[12.7]	[13.9]	[19.5]	[17.6]	[13.8]	[16.9]
Specific	80	89	85	80	90	80
	[17.3]	[16.1]	[15.1]	[19.3]	[13.8]	[16.0]
MonoF						
Common	92	92	72	71	78	81
	[11.3]	[11.3]	[17.8]	[17.5]	[23.8]	[17.1]
Specific	79	93	89	86	88	78
	[12.4]	[11.2]	[12.9]	[13.8]	[16.1]	[16.3]

Table 16

Mean reading speed in French (in ms) for the ImD, ImF and MonoF groups in Grade 3 (standard deviations in brackets)

	Frequent Words		Rare Words		Pseudowords	
	Short	Long	Short	Long	Short	Long
ImD						
Common	969	1890	1335	2134	1265	2072
	[204]	[972]	[552]	[1197]	[368]	[900]
Specific	1166	1723	1585	1989	1442	2832
	[320]	[1020]	[944]	[841]	[712]	[995]
ImF						
Common	1056	1577	1434	2045	1326	3890
	[231]	[715]	[470]	[885]	[466]	[947]
Specific	1114	1435	1347	1642	1294	2053
	[292]	[509]	[423]	[542]	[315]	[993]
MonoF						
Common	1004	1345	1240	1676	1287	1979
	[342]	[463]	[425]	[577]	[556]	[831]
Specific	991	1440	1213	1517	1227	1957
	[239]	[593]	[269]	[634]	[359]	[981]

The effect of Frequency was unreliable on accuracy, $F1(2, 124) = 21.9$, $p < .001$; $F2(2, 60) = 2.6$, $p < .10$, but was significant and reliable on latencies, $F1(2, 114) = 26.3$; $F2(2, 60) = 14.5$, both $p < .001$, as was the effect of Length (accuracy: $F1(1, 62) = 12.6$, $p = .001$; $F2 < 1$; latencies: $F1(1, 57) = 83.2$; $F2(1, 60) = 84.2$, both $p < .001$). The effect of Specificity was not significant on latencies, $F1$ and $F2 < 1$, and marginal and unreliable on accuracy, $F1(1, 62) = 5.1$, $p < .05$; $F2 < 1$. Group interacted with Frequency and with Specificity only in the analyses on latencies, but in an marginal and unreliable way, $F1 < 1$; $F2(4, 120) = 2.7$, $p < .05$, and $F1(2, 57) = 3.2$, $p < .05$; $F2 < 1$, respectively.

The interaction between Group and Length was significant on accuracy, $F1(2, 62) = 5.6$, $p < .01$; $F2(2, 120) = 3.6$, $p < .05$, and unreliable on latencies, $F1 \approx 1$; $F2(2, 120) = 4.1$, $p < .05$. On accuracy, it was due to the fact that the ImD read short items better than long items, $F1(1, 22) = 19.4$, $p < .001$; $F2(1, 70) = 3.1$, $p < .10$, whereas the ImF and the MonoF displayed equivalent performances for both materials, $F1$ and $F2 < 1$ in both groups. Coherently, no group difference was observed on latencies for the short items, $F2(2, 70) = 2.8$, $p < .10$. Still, the groups' latencies differed for long items, $F2(2, 70) = 9.9$, $p < .001$, with slower responses for the ImD and ImF than for the MonoF, $p = .001$ and $< .05$, respectively, and for the ImD than for the ImF, $p < .01$.

To summarise, in Grade 3, the ImD children's reading performances in French were close to the ones of the MonoF and ImF. However, both the ImD and the ImF read slower than the MonoF, at least long items. All three groups seem to rely on essentially the same word recognition processes. In particular, they showed comparable effects of frequency. However, it is worth noting that only the ImD showed a significant effect of length on accuracy scores.

Grade 3: Regular and irregular French word reading. Table 17 displays the mean accuracy scores and latencies in the three groups for French regular and irregular words. Both participant and item analyses were conducted on accuracy scores, including Regularity (regular vs. irregular words) in addition to the Group factor. However, only the analysis by participants was conducted on latencies, as there were too few correctly read irregular words to conduct an analysis by items.

Regularity affected performance, with regular words read more accurately, $F1(1, 62) = 775$; $F2(1, 80) = 61.1$, both $p < .001$, and more rapidly, $F(1, 26) = 4.1$, $p = .05$, than irregular words. The main effect of Group was not significant (accuracy: $F1$ and $F2 < 1$; latencies: $F < 1$), but the interaction between Group and Regularity reached significance on accuracy, $F1(2, 62) = 3.7$, $p < .05$; $F2(2, 160) = 5.1$, $p < .01$ (latencies: $F < 1$). Indeed, all three groups displayed an effect of Regularity on accuracy, but this effect was slightly stronger in the MonoF (on average, 40%, $F1(1, 11) = 496.3$; $F2(1, 94) = 51.2$; both $p < .0001$) than in the ImD (on average, 33%, $F1(1, 22) = 221.6$; $F2(1, 94) = 39.7$, both $p < .0001$) and the ImF (on average, 32%, $F1(1, 31) = 354.9$; $F2(1, 94) = 42.9$, both $p < .0001$).
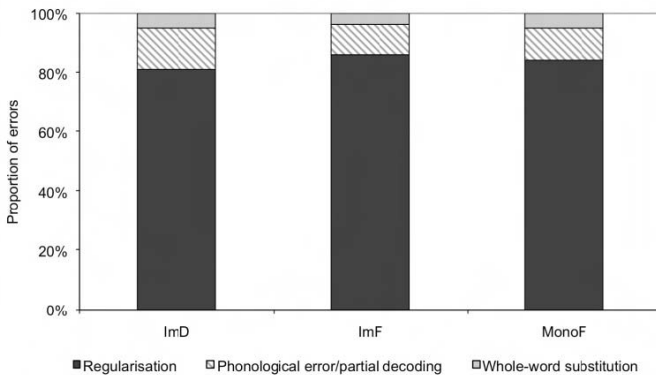
Table 17

Mean percentages of correct answers and mean reading speed (in ms) of the ImD, ImF and MonoF groups on regular and irregular words reading task in French presented in Grade 3 (standard deviations in brackets)

	Regular Words		Irregular Words	
	Accuracy scores	Mean latencies	Accuracy scores	Mean latencies
ImD	81	1327	48	1399
	[9.8]	[443]	[14.3]	[496]
ImF	82	1466	50	1607
	[8.6]	[435]	[12.3]	[652]
MonoF	84	1126	44	1211
	[5.6]	[412]	[8.4]	[533]

Errors on irregular words were classified into three categories: (i) regularisations, (ii) other phonological errors or partial decoding, (iii) whole-word substitutions. As can be seen in Figure 6, all three groups produced a very large number of regularisations, and the number of such errors formed the vast majority of the total number of errors in all three groups, $\chi^2 = 6.01$, $p > .10$.

Figure 6

Proportion (in %) of regularisation errors (in blacks), phonological errors or partial decoding (in hatched), and whole-word substitutions (in gray) for the ImD, ImF and MonoF groups in the French reading task presented in Grade 3

Grade 3: Text comprehension in French. Table 18 shows the average correct score (in %) per group. As can be seen, the three groups performed at similar levels, $F < 1$.

Table 18

Average comprehension scores (in %) in the French text comprehension task presented in Grade 3, separately for the ImF, ImD and MonoD groups (standard deviations in brackets)

	Text comprehension
ImF	94
	[7.6]
ImD	94
	[8.0]
MonoF	94
	[7.4]

Comparison between the two groups in immersion in their two languages

In order to compare the reading proficiency of the two Im groups in their two languages, ANOVAs with Language of test (French vs. Dutch) and Group (ImF vs. ImD) as factors were carried out on the words and pseudowords reading performances in both languages in Grade 2 and Grade 3.

End of Grade 2: Word and pseudoword reading in Dutch vs. French. There was a significant main effect of Language of test, which was reliable in the analyses on accuracy, $F1(1, 29) = 25.7$; $F2(1, 60) = 20.3$, both $p < .001$, and unreliable on latencies, $F1 < 1$; $F2(1, 59) = 11.9$, $p < .001$. More interestingly, the interaction between Group and Language of test was reliable in both the analyses on accuracy, $F1(1, 29) = 51.2$; $F2(1, 60) = 160.3$, both $p < .001$, and latencies, $F1(1, 14) = 4.5$, $p = .05$; $F2(1, 59) = 15.9$, $p < .001$. As shown in the top part of Figure 7, the ImF read more accurately in French than in Dutch, $F1(1, 14) = 94.9$; $F2(1, 71) = 57.4$, both $p < .001$, while the ImD read at similar levels in both languages, $F1(1, 15) = 1.8$ and $F2(1, 71) = 2.7$, both $p > .10$. In addition, as illustrated in the bottom part of Figure 7, the ImF also read somewhat faster in French than in Dutch, $F1(1, 12) = 1.9$, $p > .10$; $F2(1, 70) = 39.1$, $p < .001$. Although the ImD tended to read faster in Dutch than in French, this effect was not significant, $F1(1, 16) = 3.3$, $p > .10$; $F2 < 1$.

Grade 3: Word and pseudoword reading in Dutch vs. French. As in Grade 2, there was a significant and reliable main effect of Language of test on accuracy scores, $F1(1, 51) = 50.9$; $F2(1, 60) = 21.9$, both $p < .001$, but not on latencies, $F1(1, 28) = 3.3$, $p < .10$, $F2(1, 59) = 3.9$, $p = .05$. The interaction between Group and Language of test was highly significant and reliable in the analyses on accuracy, $F1(1, 51) = 71.6$; $F2(1, 60) = 120.1$, both $p < .001$, and significant but unreliable in the analyses on latencies, $F1(1, 28) = 2.1$, $p > .10$; $F2(1, 59) = 35.3$, $p < .001$. As shown in the top part of Figure 8, the ImF still read much better in French than in Dutch, $F1(1, 30) = 108.5$; $F2(1, 71) = 54.3$, both $p < .001$, while the ImD read at similar levels in both languages, $F1(1, 21) = 1.5$, $p > .10$; $F2 < 1$. Coherently, as illustrated in the bottom part
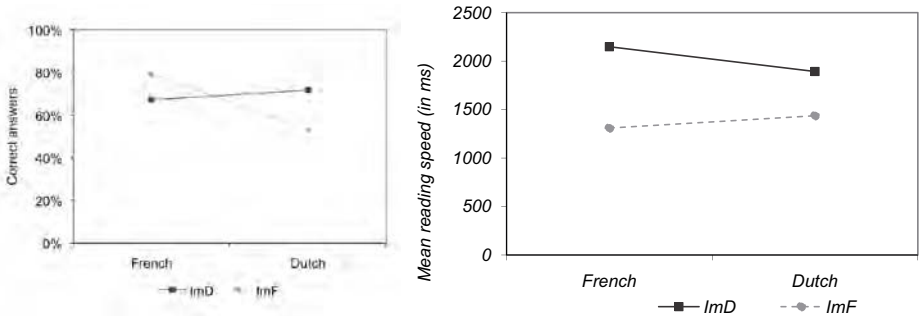
Figure 7
*Comparison of reading proficiency (top graph, in %) and reading speed
(bottom graph, in ms) of the two immersion groups (ImD: black plain line;
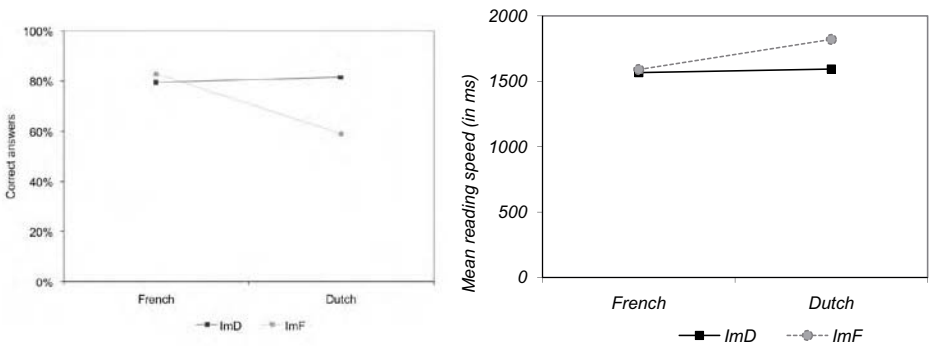ImF: gray dotted line) in their two languages at the end of Grade 2*

Figure 8
*Comparison of reading proficiency (top graph, in %) and reading speed
(bottom graph, in ms) of the two immersion groups (ImD: black plain line;
ImF: gray dotted line) in their two languages in Grade 3*

of Figure 8, the ImF read somewhat more rapidly in French than in Dutch, $F2(1, 70) = 14.3$, $p < .001$, whereas the ImD read at the same rate in both languages, $F2(1, 70) = 2.3$, $p > .10$.

Discussion

Role of oral proficiency and orthographic transparency in reading development

The first aim of the present study was to examine the relative impact of oral proficiency and of orthographic transparency on reading development in French-native children in immersion in Dutch and learning to read either in their native language, French, first (least consistent orthography) or in their second language, Dutch, first (most consistent orthography).

Previous studies have reported that children's oral proficiency in L2 contributes critically to their levels of reading in that language. For instance, Verhoeven (2000) suggested that second-language learners may have greater difficulty in recoding letter strings phonemically because they are less able to distinguish sounds in that language, and might also have difficulties in building up orthographic representations in their L2 because of the restricted size of their lexicon in that language. If this were the case, children who learn to read in a non proficient L2 first would have greater difficulties developing both the phonological and the lexical word recognition procedures than children who learn to read in their native language first. However, as reviewed in the Introduction, other studies suggest that the rate of acquisition of basic reading skills is not identical across different orthographic systems, as this development seem to be affected by orthographic transparency. In other words, in more transparent orthographic systems such as Dutch, decoding may be less demanding and GPC easier to learn and use than in less transparent orthographic systems. If this were the case, children who learn to read in a L2 with a more transparent orthographic system first would quickly attain high levels of accuracy in decoding, even though their levels of oral proficiency in their L2 is rudimentary.

Several conclusions can be drawn from the present study. First, in Grade 1, the results show that the two groups in immersion relied on different reading strategies, in accordance with the differing degree of transparency of their first language of reading instruction. The group in immersion taught to read in Dutch first (ImD) seemed to rely on the phonological procedure, while the group in immersion taught to read in French first (ImF) was more prone to rely on the lexical procedure. Indeed, like the Dutch monolinguals (MonoD), the ImD displayed a strong effect of length but a small effect of

lexicality. Moreover, these two groups produced a larger relative percentage of errors consisting in decoding the item partially or producing a pseudoword than the children instructed to read in French first.

In contrast, like the French monolinguals (MonoF), the ImF displayed a strong effect of lexicality but a small effect of length, and these two groups showed a greater tendency to produce errors consisting in whole-word substitution than the children taught to read in Dutch first.

These data extend the existing evidence for the notion that differences between languages in terms of orthographic transparency impact on the early development of reading procedures to French and Dutch, two languages differing in orthographic transparency to a much lesser extent than the pairs of languages involved in most of the previous comparisons, namely English vs. languages with very transparent orthographic systems (German, Spanish, Greek, and Welsh).

A second important finding from the present study is that, from Grade 1 to Grade 3, the Im children instructed to read in their native language first performed at the same level as the French monolinguals in tasks involving reading in French. In contrast, not surprisingly, in Grade 1 the Im children instructed to read first in their L2 performed more poorly than the Dutch monolinguals in both word and pseudoword reading tasks. However, it should be noted that for pseudoword reading they outperformed the two groups reading in French. Moreover, at the beginning of Grade 2, the Im children instructed to read in Dutch first caught up the Dutch monolinguals on pseudoword reading, even though they still performed more poorly than the Dutch monolinguals on tasks involving the reading of words or the comprehension of sentences.

This probably results from their restricted levels of lexical knowledge in that language. Indeed, at that time, the ImD did not display any effect of frequency in reading, which suggests that they still used the phonological recoding procedure even for frequent words, while the Dutch monolinguals started to use the lexical procedure to read familiar words. In addition, the reading proficiency of the ImD was highly correlated to their level of vocabulary in Dutch. These findings support Verhoeven's (2000) hypothesis that L2 learners might have difficulties in building up orthographic representations in a L2 because of the restricted size of their lexicon in that language.

A similar conclusion emerges from the results observed by the end of Grade 2 and in Grade 3. At that time, although the Im children instructed to read in Dutch first showed similar accuracy scores than the Dutch monolinguals in the reading of Dutch words and pseudowords (as well as equivalent levels of comprehension of Dutch written texts), they did not benefit from the orthographic and phonological familiarity of frequent words to the same extent as the Dutch monolinguals. This is also probably due to their poorer

levels of oral proficiency in that language.

All these results suggest that readers may not develop a frequency effect in reading words in a given language unless they have reached a high oral proficiency in that language. This might explain why neither the ImD nor the ImF showed a frequency effect when reading Dutch (because at the time of testing, their oral vocabularies did not allow frequency effects to the same extent as was the case for monolingual speakers of Dutch), while both groups displayed a strong frequency effect when reading French (their native language, in which their oral vocabularies supported a frequency effect).

However, as attested by the very good reading performance of the ImD since the end of Grade 2, our results show that learning to read in a non proficient L2 does not hamper the acquisition of GPC rules, provided that the L2 involves a transparent orthographic system. On the contrary, our findings show that when socio-economic differences between groups are controlled for, and when French-native children learning to read in Dutch first receive appropriate school support through a program of immersion, these children only show *transitory* difficulties compared to Dutch monolinguals, which are no longer observed after *only* three years of formal instruction. Indeed, at this point, the children in immersion read words and pseudowords as accurately *and* as quickly as the Dutch monolinguals, and showed similar levels of comprehension of texts.

From one orthographic system to the other…

The second aim of the present study was to examine the influence of reading acquisition in one language on the development of reading skills in the other language. An important finding is that *only* a few months after the instruction in the second language of reading acquisition began (by the end of Grade 2 or the beginning of Grade 3), the level of performance in French of the Im children taught to read in Dutch first was similar to the ones of the French monolinguals and of the Im children taught to read in French first. In contrast, the Im children in immersion who were taught to read in French first performed more poorly than both the Im children taught to read in Dutch first and the Dutch monolinguals in tasks assessing reading skills in Dutch. In other words, it took *only* a few months to the Im children taught to read in Dutch first to catch up with their monolingual peers in their native language, whereas the Im children taught to read in French first did not catch up in Dutch with the Dutch monolinguals and the Im children instructed to read in Dutch first.

This difference between the two groups of immersion cannot be entirely accounted for by the fact that the Im children taught to read in French first displayed lower levels of oral proficiency in Dutch than the children instruct-

ed to read in Dutch first, because even at the end of Grade 2, both groups of children showed much lower levels of oral proficiency, as measured in a task assessing oral sentence comprehension, than the Dutch monolinguals (respectively, 59% and 67%, vs. 81%). Moreover, despite these differences, the Im children taught to read in Dutch first caught up with the Dutch monolinguals. Rather, we hypothesise that the Im children taught to read in Dutch first had more opportunities to develop fast and efficient processes of phonological recoding thanks to the transparency of the Dutch orthographic system, and then to transfer these processes to their other language of reading instruction, than the Im children taught to read in French first.

As such, this pattern of results not only supports the hypothesis that the rather predictable GPC of the most consistent orthographic system positively influence phonological processing skills, which in turn enhance reading skills in the least consistent system, but also suggests that the learning of reading in a transparent orthographic system allows the children to literally train and "over learn" phonological recoding skills to such an extent that it takes them only a few months to subsequently learn to read in another language and catch up with children who were always instructed to read in that language. In other words, transparent orthographic systems could really constitute training tools for the development of fast and effective phonological recoding skills, compared to less transparent or opaque orthographic system.

It is also worth mentioning that no evidence of negative transfer from Dutch to French was observed. Indeed, the ImD did not display a greater effect of regularity than the French monolingual or the ImF. In addition, all three groups made a large number of errors consisting in regularisations, and these formed the vast majority of the total number of errors across the three groups. This may seem surprising. Indeed, in Grade 1 a difference in reading procedures for Dutch and French readers had been attested by their different susceptibility to lexicality and length. If the ImD and MonoD readers were making use of the phonological assembly procedure that is promoted by Dutch orthography, they should make more regularisation errors than the ImF and MonoF readers. Still, a pure phonological assembly procedure is probably not characteristic of more mature Dutch readers (de Jong, 2006; Martens & de Jong, 2008). Although Dutch orthography is highly regular, several deviations from a one-to-one correspondence occur. Therefore, at least some (although a few, in comparison to French) orthographical forms, rules and exceptions have to be memorised by Dutch learners, and some words cannot be read adequately using only the phonological route (Verhoeven, Baayen, & Schreuder, 2004). One would thus expect the children who learn to read in Dutch to progressively start using the lexical procedure to read familiar words. This was actually the case for the MonoD already at

the beginning of Grade 2, when they showed a weaker effect of length for the frequent words than for the rare words and pseudowords. And the ImD children, who did not display any lexicality effect at all in Grade 1, did show such an effect from the end of Grade 2 on, even though they still experienced difficulties with building up a L2 reading lexicon at the beginning of Grade 2 and did not benefit to the same extent as Dutch monolinguals from the effect of word familiarity in reading even in Grade3.

The fact that the Dutch orthographic system, although much more transparent than the French one, still leads to developing the lexical procedure (at least after some reading experience) may also explain why our results are inconsistent with those reported by Mumtaz and Humphreys (2001). As a matter of fact, in that study, the negative transfer from the most consistent orthographic system (Urdu) to the least consistent one (English) was attributed to the greater reliance on non-lexical processing by bilingual children when reading irregular words in English, a procedure which led them to commit many regularisation errors. This discrepancy may be accounted for if the contrast between French and Dutch in terms of orthographic transparency were less important than the contrast between English and Urdu. Indeed, Urdu orthography strongly emphasises phonological rather than visual orthographic strategies, since the written symbols in Urdu vary visual forms depending on their position in a word (Mumtaz & Humphreys, 2001). Another explanation may be that in Mumtaz and Humphreys' study, English was the children's L2, and it is therefore difficult to separate the general effect of bilingualism from negative transfer.

The degree and the speed with which an orthographic system pushes beginning readers to develop an orthographic lexicon in addition to the phonological procedure is thus probably crucial in understanding both positive and negative transfers from one orthographic system to the other.

Conclusions

Previous studies have outlined the importance of developing children's language proficiency before exposure to literacy in a second language. The results of our study as well as of other studies in different linguistic contexts cast some doubt on the pervasive role that oral proficiency might play in the development of basic word recognition skills in a second language. More specifically, the results of the present study suggest that there are potentially significant benefits to learn to read first in the most consistent orthographic system, even when it is the least proficient language, since orthographic consistency positively influences phonological recoding skills, and therefore the development of fast and accurate mechanisms of word identification, which can be then transferred across the two languages.

If this were the case, then, we would predict that French-speaking children in immersion in a language displaying an opaque orthographic system, such as English, would display the opposite pattern of results as the one we have observed in the present study. Indeed, it can be assumed that when the L2 presents the least consistent orthographic system, it would constitute a greater difficulty for learning to read, since the orthographic irregularity would unable children to decode unfamiliar words with accuracy by relying on the phonological procedure. Specifically, French-native children in immersion taught to read in French first would catch up English monolinguals *faster* than French-native children in immersion instructed to read in English first. Because immersion programs in Belgium involve French and Dutch as well as French and English, it would be extremely interesting if further studies were conducted to examine this prediction.

Indeed, if this prediction was supported by empirical data, this would suggest that in the context of the development of reading in two (or more) language(s), reading tuition should take place in the most transparent orthographic system first in every case, because the gains from developing *and* training fast and accurate phonological recoding skills thanks to the transparency of the system, and then to transfer these recoding skills to the other(s) language(s), would override the gains from teaching reading in the language in which the child has the relatively greater level of oral proficiency.

If this were the case, the suggestion that some orthographic systems should be reformed in order to provide more consistent grapho-phonological correspondences would receive scientific support, to the extent that more transparent orthographic systems provide children with greater opportunities to develop fast and efficient phonological recoding skills, which are essential to the successful development of reading in all alphabetic languages, and which could be rapidly transferred across language in children learning to read in more than one language. This important issue should be carefully addressed with further studies examining reading development in different pairs of languages differing in terms of orthographic consistency.

References

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

Bos, M., & Reitsma, P. (2003). Experienced teachers' expectations about the potential effectiveness of spelling exercises. *Annals of Dyslexia, 53*, 104-127.

Burnage, G. (1990). *CELEX: A guide for users.* Nijmegen: Center for Lexical Information.

Carlisle, J., & Beeman, M. (2000). The effects of language of instruction on the reading and writing achievement of first-grade Hispanic children. *Scientific Studies of Reading, 4*, 331-353.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope – An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments & Computers, 25*, 257-271.

Content, A., Mousty, P., & Radeau, M. (1990). BRULEX: Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90*, 551-566.

Da Fontoura, H., & Siegel, L. (1995). Reading, syntactic, and working memory skills of bilingual Portuguese English Canadian children. *Reading and Writing: An Interdisciplinary Journal, 7*, 139-153.

D'Angiulli, A., Siegel, L., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics, 22*, 479-507.

Defior, S., Martos, F., & Cary, L. (2002). Differences in reading acquisition development in two shallow orthographies: Portuguese and Spanish. *Applied Psycholinguistics, 23*, 135-148.

de Jong, P.F. (2006). Units and routes of reading in Dutch. *Developmental Science, 9*, 441-442.

Dunn, L.M., Thériault-Whalen, C., & Dunn, L.M. (1993). *Échelle de vocabulaire en images peabody*. Toronto: Psycan.

Ellis, N., & Hooper, M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics, 22*, 571-599.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies in Reading, 2*, 31-54.

Galland, J. (1941). A study of the structural elements of the French language. *The Modern Language Journal, 25*, 477-487.

Geva, E., & Siegel, L. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12*, 1-30.

Goetry, V. (2002). *Étude longitudinale comparative du développement des compétences métaphonologiques et orthographiques d'enfants bilingues scolarisés en langue seconde*. Unpublished doctoral dissertation, Université Libre de Bruxelles.

Goswami, U., Gombert, J., & de Barrera, L. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19*, 19-52.

Goswami, U., Porpodas, C., & Wheelwright S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education, 12*, 273-292.

Goswami, U., Ziegler, J., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics, 24*, 235-247.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia, 9*, 317-323.

Hutzler, F., Ziegler, J., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*, 273-296.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47*, 145-171.

Lambert, E., & Chesnet, D. (2001). NOVLEX: Une base de données lexicales pour les élèves de primaire. *L'Année Psychologique, 101*, 277-288.

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education, 15*, 239-257.

Lange, M., & Content, A. (1999). *The grapho-phonological system of written French: Statistical analysis and empirical validation*. Paper accepted for ACL®99: the 37th Annual Meeting of the Association for Computational Linguistics. Thematic session on Computational Psycholinguistics, University of Maryland, June 20-26.

Lecocq, P. (1996). *L'E.C.O.S.S.E.: Une épreuve de compréhension syntaxico sémantique*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Martens, V.E.G., & de Jong, P.F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading, 31*, 40-54.

Mumtaz, S., & Humphreys, G. (2001). The effects of bilingualism on learning to read English: Evidence from the contrast between urdu-English bilingual and English monolingual children. *Journal of Research in Reading, 24*, 113-134.

Nunn, A.M. (1998). *Dutch orthography: A systematic investigation of the spelling of Dutch words*. The Hague, The Netherlands: Holland Academic Graphics.

Plaut, D., McClelland, J., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 94*, 523-568.

Rastle, K., & Davis, M.H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 307-314.

Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen. Derde herziene versie gebaseerd op nieuw onderzoek in Nederland en België*. Lisse, Switzerland: Swets & Zeitlinger.

Schutzenberg, A., & Mavré, D. (1981). *Traduction du manuel Progressive matrices de J. Raven*. Issy-les-Moulineaux: Éditions Scientifiques et Psychologiques.

Sebastien-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non native phonemic contrasts in early bilinguals. *Cognition, 72*, 111-123.

Seymour, P., Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in european orthographies. *British Journal of Psychology, 94*, 143-174.

Share, D.L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151-218.

Share, D.L., & Stanovich, K.E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology, 1*, 1-57.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Spencer, L., & Hanley, J. (2003). The effects of orthographic consistency on reading development and phonological awareness; evidence from children learning to

read in Wales. *British Journal of Psychology, 94*, 1-28.

Spencer, L., & Hanley, J. (2004). Learning a transparent orthography at five years old: Reading development of children during their first year of formal reading instruction in Wales. *Journal of Research in reading, 27*, 1-14.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition, 100*, 369-388.

Tranel, B. (1987). *The sounds of French*. New-York: Cambridge University Press.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4*, 313-330.

Verhoeven, L., Baayen, R.H., & Schreuder, R. (2004). Orthographic constraints and frequency effects in complex word identification. *Special issue of Written Language & Literacy, 7*, 49-59.

Véronis, J. (1986). Étude quantitative sur le système graphique et phono-graphique du français. *European Bulletin of Cognitive Psychology, 6*, 501-553.

Wade-Woolley, L., & Geva, E. (2000). Processing novel phonemic contrasts in the acquisition of L2 word reading. *Scientific Studies of Reading, 4*, 295-311.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition, 51*, 91-103.

Yvon, F., de Mareuil, P.B., d'Alessandro, C., Auberge, V., Bagein, M., Bailly, G. et al. (1998). Objective evaluation of grapheme to phoneme conversion for text-to-speech synthesis in French. *Computer Speech and Language, 12*, 393-410.

Ziegler, J., Jacobs, A., & Stone, G. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, and Computers, 28*, 504-515.

Ziegler, J., Montant, M., & Jacobs, A.M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*, 533-554.

Ziegler, J., Stone, G., & Jacobs, A. (1997). What's the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, and Computers, 29*, 600-618.

Zonneveld, W. (1978). *A formal theory of exceptions in generative phonology*. Lisse: Peter de Ridder Press.