The performance of L2 French children on the LITMUS-QU Nonword repetition task during their first year of exposure: impact of age, vocabulary size, verbal-short term memory and phonological awareness

Letícia Almeida[1] and Christophe Coupé[2]

[1]Center of Linguistics, School of Arts and Humanities, University of Lisbon (Lisbon, Portugal) and [2]Department of Linguistics, The University of Hong Kong (Hong Kong SAR, China)

CAMBRIDGE UNIVERSITY PRESS

Abstract

In this study, we describe the performance of 62 newly immigrated children to France at a nonword repetition task (LITMUS-QU-NWR-FR) designed to evaluate bilingual children's syllable structure. Children were between 6;0 and 9;1 and had diverse language backgrounds. They participated in our study during their first year of exposure to French. The majority of our children exhibited a good performance on the task. The variation observed is related to: (i) the properties of the nonwords: items with complex syllables are more difficult, as are items with three syllables in length; (ii) phonological awareness: children with a more developed L2 phonological awareness perform better at the task; and (iii) receptive vocabulary size: children with a larger L2 vocabulary size perform better at the task. Overall, our findings provide support for the argument that the LITMUS-QU-NWR-FR task can be used shortly after the onset of exposure to the L2.

Introduction

Nonword repetition (NWR) tasks are commonly used for clinical assessment in various languages. They have recently been pointed out to be good tools for evaluating both simultaneous and successive bilingual children because they allow us to differentiate between typically developing children and children with developmental language disorders (Almeida et al., 2017; Paradis, 2016; Schwob et al., 2021). The main reason for that is that they rely less on previous lexical knowledge than other tasks.

Although NWR tasks rely on phonology, most of those tasks evaluate verbal short-term memory as they manipulate word length and include nonwords up to five or seven syllables (Gathercole et al., 2007). In such cases, it is difficult to know if children have poor results at the whole task due to poor verbal short-term memory or to difficulties related to phonological competence. This is especially true if the task does not control for phonological complexity equally across word length (Gallon et al., 2007). For example, Cilibrasi et al. (2018) showed that phonological complexity, in particular the presence of a cluster, negatively influenced the target repetition of the nonwords. As those clusters only appear in the four- and five-syllable nonwords of the Children's Test of Nonword Repetition, these results could be an effect of either phonological complexity or word length (i.e., verbal short-term memory), or of both.

Among the NWR tasks that take into account bilingual contexts in their construction, the LITMUS (*Language Impairment Testing in Multilingual Settings*)-QU (*Quasi Universal*)-NWR task (dos Santos & Ferré, 2018, for the French version) aims to assess children's phonology through phonological complexity. This is in order not to penalize bilingual learners because of their shorter exposure to the test language. Additionally, it relies less on verbal short-term memory because the items are shorter than in other tasks (dos Santos & Ferré, 2018). This task, which has been adapted in various languages, has proven to be effective in distinguishing both monolingual and bilingual children with typical development from those with developmental language disorder. Studies have been conducted in French, German, and Lebanese Arabic (Abi-Aad & Atallah, 2020; Almeida et al., 2017; Grimm, 2022; Hamann & Abed Ibrahim, 2017). Looking at the overall accuracy at the task, several studies show that typically developing bilingual children reach the same score as monolinguals (Almeida et al., 2017; Grimm, 2022). A few studies on LITMUS-QU-NWR take additional factors into account. Almeida et al. (2017) showed no influence of variables related to bilingualism in children from three different language backgrounds (Arabic, Portuguese and Turkish), such as quantity of input or Length of Exposure (LoE), on the accuracy at the task. It must be noted, though, that all the children, simultaneous and successive bilinguals, had more

than a year of exposure to the target language. Chilla et al. (2021) showed no first language (L1) effect on the overall accuracy at the task, but they looked at children with only three different L1s (Arabic, Portuguese and Turkish). Grimm (2022) and Grimm, A. & Domahs (2023) expanded the result on insensitivity of the task to children's L1s, showing that eL2 learners of German from 12 different L1s overall exhibited a good accuracy at the task with no L1 effect.

Almeida et al. (2017) and dos Santos and Ferré (2018), that are among the first studies conducted on the LITMUS task, included simultaneous and successive bilinguals in their analysis. Considering both categories together possibly hid potential effects of LoE and L1 in successive bilinguals, as they have less exposure to their L2 comparing to simultaneous bilinguals. Recently, some studies have focused exclusively on successive bilinguals but, even in these cases, children have typically been exposed to their L2 for several years (Abed Ibrahim et al., 2020; Grimm, 2022). The only exception, to our knowledge, is Scheidnes (2020) who focused on L1 English successive bilinguals being exposed to French for several months only. She found that, like monolinguals, bilingual children reached 80% of accuracy at the task even with little exposure. The contribution of this study is valuable because it is the first to show that children with very little exposure to their L2 have an overall good performance on the LITMUS-QU-NWR task. It is difficult, however, to assess whether this result would stand with children having L1s with different phonological properties. The implications of the studies conducted so far on the LITMUS-QU-NWR tasks are that there is little or no evidence that variables related to bilingualism, such as quantity and quality of the input or LoE, influence the outcome. The same can be said of L1 effects in simultaneous and successive bilinguals, even in the case of very little exposure to the L2. That makes it a very good task for evaluating phonology in an heterogenous population of bilingual children.

The present study aims to fill a gap in the LITMUS-QU-NWR literature concerning the impact of having diverse L1s during the first stages of exposure to an L2. We focus exclusively on the very first months of exposure to French L2 during childhood (like Scheidnes, 2020), in a context of immigration to France where children have very diverse L1 backgrounds (like Grimm, 2022). Our study therefore reflects the current reality of immigration in France, and we want to address the issue of assessing phonology in this specific population, which is a huge challenge for speech therapists (Armon-Lotem & Grohmann, 2021). Previous studies investigating the development of early second language (eL2) children's phonology suggest that they develop a solid grasp within a year of exposure to the L2 (Duncan & Paradis, 2016; Tessier et al., 2013). This is particularly important for the field of clinical linguistics, as it points to the possibility of detecting phonological impairment prior to a year of exposure. To do that, though, we still need to better understand what influences eL2 children's performance on the LITMUS-QU-NWR task. Our present study stands out by investigating the impact of both phonological and non–phonological factors on children's NWR accuracy. In the next subsections, we discuss variables that may influence performance on NWR tasks.

Length of Exposure

Duncan and Paradis (2016) found that LoE positively influenced eL2 children's performance on a standardized NWR task: the Comprehensive Test of Phonological Processing (Wagner et al., 1999). The authors noted, however, that their measure of LoE rests only on input quantity and does not consider input quality. On the contrary,

Mezziane and MacLeod (2017) did not find any correlation between 'Percentage of Consonants Correct' and the use of French at home in their population of eL2 French learners. This contradictory finding might be explained by methodological aspects: Duncan and Paradis (2016) focused on a NWR task whereas Mezziane and MacLeod used spontaneous productions. Scheidnes (2020) focused on early English learners of French with 17 to 23 months of exposure through immersion schooling, aged 6;4 to 7;5. The cumulative exposure to French was only 2 to 4 months. Yet, the children performed on average above 80% of target productions on the LITMUS-QU-NWR task, leading the author to conclude that LoE had low impact on the performance of English children learning French as an L2. One should note, however, that this study has a limited range in LoE, which hinders the detection of effects. Additionally, the children examined in Duncan & Paradis (2016) were younger than the ones analyzed in Scheidnes (2020). In this study, we aim at further testing the impact of LoE on a NWR task.

Phonological complexity

Children acquire the sounds of their languages according to the position they occupy within the syllable. This position is traditionally described in terms of syllable constituents: onset, nucleus, coda. Vowels are part of the nucleus whereas consonants can either be syllabified as onsets or codas. Onsets are pre-vocalic and can either be singleton (ex: _pas_, "not") or complex (branching), as in _plat_ ("course"). Codas are post-vocalic and generally singleton (ex: _partie_, "part") (Selkirk, 1982). Final consonants have a special status because they can either be syllabified as codas or as onsets of empty-headed syllables depending on the target language, the latter option being the mainstream analysis for French (Piggott, 1999). French allows four combinations of an obstruent followed by a liquid in branching onset position: a plosive + a rhotic (e.g., [pʁ]_incesse_, "princess"), a plosive + a lateral (e.g., [pl]_ein_, "full"), a fricative + a rhotic (e.g., [fʁ]_uit_, "fruit") and a fricative + a lateral (e.g., [fl]_eur_, "flower"). Additionally, an initial /s/ can precede such clusters, as in strict /stʁikt/ 'strict'. It is virtually unrestricted concerning its syllable-final consonant inventory, either word-internally or word-finally: almost all consonants can appear in this position (e.g. be[l] 'beautiful' or a[l]bum 'album' (Dell, 1995). Typically, during L1 acquisition, singleton onsets are acquired before branching onsets, codas, and final consonants (Kehoe & Havy, 2019; Rose, 2000, dos Santos, 2007). Consequently, these last three structures are traditionally viewed as complex during monolingual acquisition.

Previous studies have shown that children, including eL2 learners, have trouble acquiring complex phonological structures like branching onsets (Tessier et al., 2013) and codas (Duncan & Paradis, 2016; Morrow et al., 2014; Rattanasone & Demuth, 2014). Meziane and MacLeod (2017), who evaluated eL2 learners of French in Montreal with 12 different L1s, also found that the children had a lower 'Percentage of Correct Consonants' word-finally. Scheidnes (2020) also reported that eL2 French learners performed worse on nonwords with /l/ in coda position and clusters in general, suggesting that the phonological complexity of the nonwords may decrease accuracy at a NWR task, even for children who have a good overall performance. Grimm (2022) found that eL2 learners of German had lower performances than their monolingual peers on the complex structures of the LITMUS-QU-NWR task. Taken together, these studies highlight difficulties with the same complex phonological structures as in monolingual acquisition (branching onsets, codas and final consonants). This finding is robust: it appears for children

learning different L2s (English, French and German), with different age ranges (from 3 to 9;11), and whether based on spontaneous or elicited productions (NWR or picture-naming task). The implication is that complex structures pose difficulty for virtually all children in different learning contexts (monolingual and bilingual), with different ages and from diverse languages. For this reason, we will focus on these structures in our study.

L1 influence

Many studies conducted on bilingual children have reported cases of cross-linguistic interaction between the children's languages (Kehoe, 2015). Whereas most of those studies are based on simultaneous bilinguals and focus on cross-linguistic interaction while the two languages are being acquired simultaneously, some studies report cases of L1 influence for eL2 children. The main difference between these children and simultaneous bilinguals is that, in principle, eL2 children have acquired (most of) their L1 phonological system, so we expect influence of their L1 on their L2. Rattanasone and Demuth (2014) focused on three-year-old children with Mandarin as L1 and learning English as an L2. Children had between 6 to 21 months of exposure to English. Using an elicited imitation task, the authors noticed some difficulties with the target production of (final) codas in English, which could be explained by the restrictions on codas in Mandarin. Duncan and Paradis (2016) made a similar observation using a NWR task. They noted that overall, the children produced word-medial and word-final codas significantly less accurately than onsets, and that codas did not reach 80% of target production. Children's L1, however, affected this pattern, as children with a South Asian L1 were more accurate on codas than those with a Chinese L1. Grimm & Domahs (2023) also found L1 influence in the data of 26 eL2 learners of German but in specific structures only: while word-initial clusters were acquired and showed no sign of L1 influence, the accuracy of word-final clusters was impacted by their properties in the L1 of the children. Tessier et al. (2013) concluded that their eL2 English learners did not master branching onsets, since most of the children exhibited less than 80% of target production during their first year of exposure to English. They too noticed that accuracy was influenced by L1: children with a South Asian background were less accurate at producing branching onsets than children with a Chinese background. Thus, all these studies suggest that L1 properties influence eL2 children's performance on their L2, in the sense that the presence of a specific property in the L1 facilitates its production in the L2. This result holds for young and older eL2 learners, and for spontaneous and elicited productions.

Lexical knowledge

The relationship between eL2 phonological development and vocabulary size is unclear. Monolingual children with good phonological representations tend to have a greater vocabulary size (Stoel-Gammon, 2011), like simultaneous bilinguals (Kehoe & Havy, 2019). To our knowledge, only three studies have examined this link for eL2 children, since most of the studies that have explored this topic were conducted on simultaneous bilinguals (e.g., Kehoe & Havy, 2019). In their study, while controlling for other factors, Duncan and Paradis (2016) found a positive effect of L2 (English) receptive vocabulary size on the performance on a NWR task: children, aged between 5;0 and 6;9, with larger vocabularies (assessed with the Peabody Picture

Vocabulary Test-IIIR) had more accurate productions. Meziane and MacLeod (2017) did not find such an effect on 6-year-old eL2 French children. After evaluating children's phonological performance with a picture elicitation task and calculating their Percentage of Correct Consonants, they found no correlation between this measure and the size of their receptive vocabulary (also assessed with the Peabody Picture Vocabulary Test). Finally, Tessier et al. (2013) did not find any influence of receptive vocabulary (assessed with the Peabody Picture Vocabulary Test) or expressive vocabulary (evaluated through spontaneous speech) on the performance on branching onsets in eL2 learners of English, aged 5;4 to 6;9. In sum, findings about the relationship between phonological performance and vocabulary size are mixed. Although receptive vocabulary was assessed with the same tool in all three studies, this was not the case for phonological performance. It is thus possible that the effect of receptive vocabulary size impacts the performance on NWR but not on other tasks. In this study, we aim at further testing the impact of receptive vocabulary size on a NWR task.

Age

To our knowledge, only two studies have examined the influence of eL2 children's age on their phonological performance. Schwob and Skoruppa (2022) studied 30 French-Portuguese bilingual children, aged 5 to 7, 12 of which were successive bilinguals (i.e., eL2). The authors found that children's performance on three different NWR tasks increased with age but did not discuss this specific result. Duncan and Paradis (2016) also found that older eL2 children had a better performance on the NWR task, independently of their LoE or of their L2 receptive vocabulary size. This is interpreted by the authors as being the consequence of the maturation of the verbal short-term memory, based on the argument that in eL2 population, age is independent of linguistic experience.

Phonological awareness

The positive impact of good phonological awareness skills is well established for L1 and L2 reading (Hogan et al., 2005; Genesee and Geva, 2006). Its positive impact on foreign word learning in a population of eL2 children has also been shown (Hu, 2003), suggesting that phonological awareness can also benefit oral language and not just reading proficiency. Engel de Abreu and Gathercole (2012) noted that L1 phonological awareness skills positively impacted L2 sound structure acquisition in eL2 children, highlighting that phonological awareness can impact L2 phonology and not just vocabulary. To our knowledge, the influence of L2 phonological awareness on L2 phonological performance has not been investigated so far for eL2 children. If L1 phonological awareness skills promote L2 phonological acquisition, it is not unreasonable to think that good L2 phonological awareness skills will have the same impact. We want to explore this hypothesis in the present study.

Current study

Even if the LITMUS-QU-NWR task has been identified as a relevant tool to identify children with and without developmental language disorder, independently of bilingualism (Almeida et al., 2017, Grimm, 2022, among others), there are still several gaps in the literature. First of all, we do not know if the cut-off (80% of accuracy) established in Almeida et al. (2017) for the French version of the task and derived from the comparison of four groups of

children stands for successive bilinguals with little exposure to the L2. Typically, previous studies incorporated both simultaneous and successive bilinguals. The studies that have focused so far exclusively on successive bilinguals include children with more than a year of exposure to the target language (e.g. Grimm, 2022), or children with a shared L1 (Scheidnes, 2020). Additionally, most of the eL2 children included in previous studies are exposed to their L2 after 2;00 (Grimm, 2022) and 4;4 (Duncan & Paradis, 2016), i.e., during a period in which their L1 phonology is not fully acquired.

In the present study, we focus on school-age children, exposed to their L2 after the age of 6. This is motivated by two reasons: in principle, by that age, their L1 phonology is close to being fully acquired, even if some variation may persist. We also want to capture a situation in which they potentially could be identified with a developmental language disorder in France, where bilingual children are typically diagnosed from the age of six (Almeida et al., 2017). We restrict the LoE of the participants to a maximum of 11 months to specifically capture the early stages of exposure to the L2 and test whether children at this stage show a good performance on the task. We include children with diverse L1s since the effect of the L1 on the performance on the LITMUS-NWR task has been assessed only with three languages (Arabic, Portuguese and Turkish) (Grimm, 2022, who focused on 12 different L1s, did not take this variable into account in her analysis). Further, a surprising gap in the LITMUS-QU-NWR literature is the absence of analysis of individual variation among children. This is probably because the task reaches high scores, thus reducing the scope of variation. Nevertheless, Duncan and Paradis (2016) noted several external factors that affect eL2 children performance on a NWR task, namely vocabulary size, verbal short-term memory, age and LoE. In this study, we consider the same predictors to investigate if the previous results hold with the LITMUS-QU-NWR task. We also expand the approach by assessing the potential effect of phonological awareness, a factor which has, to our knowledge, not been investigated so far.

Based on the literature, our research questions are as follows:

(1) Will L2 children reach 80% of correct repetition at the LITMUS-QU-NWR task shortly after having been exposed to French? 80% of correct repetition is the cut-off score previously established for this particular task, which allows us to separate children with and without developmental language disorder, independently of bilingualism, with good levels of specificity and sensitivity (Almeida et al., 2017). As English French learners reached this percentage after only several months of exposure (Scheidnes, 2020), we expected that will be the case in the current study.

(2) How will L2 children perform on complex phonological structures, namely branching onsets, syllable-final consonants (word-medially and word-finally) and #sC clusters? L2 Children have been reported to exhibit difficulties with these specific structures even when their L2 phonological system is good overall (Scheidnes, 2020). For that reason, we expect the performance on these structures to be below 80% even if their overall performance is greater than that.

(3) How do phonological complexity, LoE, age, vocabulary size, verbal short-term memory and phonological awareness influence children's overall performance on the NWR task? Given previous results, we expect an inhibitory effect of phonological complexity (Scheidnes, 2020) and a facilatory effect of LoE, age, vocabulary size, verbal short-term memory, L1 syllable

complexity (Paradis & Duncan, 2016) and phonological awareness (Engel de Abreu and Gathercole, 2012).

Methods

Ethical statement

This research and all its procedures were approved by the Ethics Committee of the School of Arts and Humanities at the University of Lisbon (Lisbon, Portugal) (reference number 12_CEI2022). We obtained parental written consent for all the children included in this study after informing the families about the goals of the research, their right to withdraw at any time without consequences, and the respect of their privacy and anonymity.

Participants

In France, all children are automatically registered at public schools at their arrival in the country. They attend normal schooling but are enrolled in a special program during their first year in primary school called *Unité Pédagogique pour l'Enseignement des Allophones Arrivants* (pedagogical unit for the teaching of newly arrived allophones), that they attend several hours per week. We recruited the children of the present study through the teachers of this specific unit in Lyon. We recruited 73 children in total but excluded those who had more than 12 months of exposure to French or were exposed to French prior to their arrival in the country. We eventually included 62 children who had between 1 and 11 months of exposure to French ($M = 5.44$; $SD = 2.17$) and were aged between 6;0 and 9;1 ($M = 89.63$ months; $SD = 8.63$). These children can be classified as early L2 learners, as they are exposed to their L2 after the age of 3 (Meisel, 2018). They had 18 different L1 backgrounds: some of them where monolingual prior to their arrival to France, speaking either Arabic ($n = 9$), European Portuguese ($n = 7$), Albanese ($n = 6$), Italian ($n = 5$), Romanian ($n = 5$), Brazilian Portuguese ($n = 4$), Armenian ($n = 3$), Japanese ($n = 2$), Mandarin ($n = 1$), Kabyle ($n = 1$), Spanish ($n = 1$) or Russian ($n = 1$), while some were already bilingual in Italian and Arabic ($n = 12$), Spanish and Arabic ($n = 1$), Russian and Ukrainian ($n = 1$), Russian and Armenian ($n = 1$), Armenian and German ($n = 1$) or yet Russian and Chechen ($n = 1$). Detailed information for each child – gender, age, L1, vocabulary score, phonological awareness score and verbal short-term memory score – is presented in the supplementary materials, Table S1.

In the field of bilingual acquisition, a challenge is to account for the variation observed in the children's productions as it may be due to differences in factors related to bilingualism such as LoE, dominance, quantity and quality of exposure to a particular language. In this paper, we tried to avoid as much as possible the impact of these external variables by choosing a group of children quite homogeneous in terms of quantity and quality of exposure to French: all were immigrated children with no previous contact with French, exposed to French only at school since their arrival, and speaking in their L1 with their families. In other words, for all the children, French was the weak language, as they had started their contact with it only several months before their participation in our study.

It was not possible to conduct an in-person interview with all the parents, but they completed a written questionnaire, either translated into their native language or in English, aiming at gathering information about the child's exposure to French before arrival in the country, exposure to other languages, date of arrival in France and birthdate. Most of this information was obtained with the help

of the specialized French teachers. We did not recruit children who had been referred to a psychological evaluation or presented huge difficulties learning to read and write, in order to avoid cases of atypical language development.

Protocol

The first author assessed all the children with an experimental task evaluating phonology (the LITMUS-QU-NWR task) and standardized French tasks assessing receptive vocabulary, verbal short-term memory and phonological awareness. The entire session took around 30 minutes per child and occurred at school, in a separate-room.

The standardized tasks

Children's receptive vocabulary was assessed using a subtest of the *Nouvelles Epreuves pour l'Examen du Langage* battery (Chevrie-Muller & Plaza, 2001).[1] Concretely, children are presented with six sets of eight pictures and for each set are asked to show the six pictures that correspond to six words elicited by the experimenter. In total, the task thus consists in identifying 36 words. Raw scores correspond to the total of correct answers (i.e., words identified correctly). It is also possible to calculate z-scores based on the results of French monolingual children. However, for the present study, we relied on raw scores because children with so little exposure to French cannot be directly compared to children who grew up in an exclusively French environment. Moreover, as explained below, we included age as a fixed factor in our statistical analysis. Verbal short-term memory was assessed using a subtest of the Exalang battery (Thibault et al., 2003; Thibault et al., 2012). This task is computerized with pre-recorded instructions and stimuli. Children are asked to repeat several sequences of numbers. If the child fails, a new sequence with the same number of items is presented. If the child fails twice, the task stops. Raw scores are between 3 and 7. Once again, we only considered raw scores for this task and not z-scores, for the same reasons as above.[2] Phonological awareness was also assessed using a subtest of the Exalang battery. Children completed two different tasks according to their age, as specified in this battery: children between 6;0 and 7;11 performed a phoneme inversion task (raw scores from 0 to 6) and children between 8;0 and 9;1 performed a task involving different manipulations: phoneme inversion, phoneme substitution, phoneme counting, phoneme deletion and rhyme-providing[3] (raw scores between 0 and 11). As the two tasks are different, we chose this time to use z-scores to merge them and consider a unified measure of phonological awareness. Five children were not able to complete this task.

The NWR task

We tested the children with the LITMUS-QU-NWR-FR task (dos Santos & Ferré, 2018). This task was specifically conceived to evaluate children's L2 phonology, especially regarding syllable complexity, with the clinical aim of disentangling language impairment and typical development in this population. For that reason, it includes nonwords with structures that are largely attested in the world's languages. Concretely, this tool elicits the repetition of 71 nonwords made of 3 vowels (/a, i, u/) and 5 consonants (/p, k, f, s, l/) that combine in different syllable types: CV, CCV, CVC, #sCV[4], #sCVC, #sCCV, Cs# and sC#. Nonwords vary between one and three syllables in length to minimize the effect of working memory on the production of nonwords (see dos Santos & Ferré, 2018 for detailed explanations about the creation of this tool). Two additional nonwords were elicited at the beginning of the task for training. The task was presented to the child in the form of a PowerPoint presentation with pre-recorded oral stimuli. Children were told that an alien would teach them their own language and that they should repeat exactly what they heard.

Data transcription and criteria for correct repetition of nonwords

Children's productions at the NWR task were audio-recorded and transcribed by the first author using Phon (Hedlund & Rose, 2020; Rose et al., 2006), a program specifically designed to segment, transcribe, and query phonological data. Whenever the first author was unsure about the phonetic transcription of the data, another trained linguist reviewed the transcription until agreement was attained. This was the case for 8% of the corpus. We then calculated the children's raw score at the task and percentage of correct repetition. Items were coded as 1 when the entire item was produced target-like, and as 0 each time there was at least one error in the production of the item. Nevertheless, following previous work based on the LITMUS-QU-NWR-FR task, we did not consider errors that do not lead to a phonemic contrast in French because the goal of this tool is to be used in clinical settings where extensive qualitative analyses are not often carried out. Therefore, we only considered vowel errors if the substitution induced a phonemic change in French. Errors on stress assignment were not considered (there were very few occurrences of such errors in our corpus anyway). Following the coding convention for the LITMUS-QU-NWR tasks, we did not consider voicing errors either, even if they create a phonemic contrast in French, since the contrast between voiced and voiceless consonants is particularly difficult to perceive, mainly due to differences in Voicing Onset Time duration across languages. Therefore, an item that was target-like except for voicing (e.g., [piku] > [biku]) was considered as correct.

For branching onsets, we considered as target-like the production of the two target segments. Substitution of at least one segment, as well as deletion or epenthesis, were considered non-target. For lateral codas, we considered as correct the production of a lateral even if it was produced as a velar lateral (e.g., [pilfu] > [piɫfu]), because this did not create a phonemic opposition. All other productions were considered non-target. For the final lateral, we checked for vowel insertion using Praat (Boersma & Weenink, 2019) and we considered that a vowel was inserted only when we could observe a variation of formant trajectory and a change in the temporal envelope. This enabled us not to penalize the children when they produced a final release. The cases of (true) vowel insertion were considered as erroneous.

---

[1]Expressive vocabulary was not considered a measure because we focused on children with very little exposure to French so that their expressive vocabulary was quite poor. We did not want to put the children in situations in which they would feel failure at a task.

[2]As an anonymous reviewer pointed out, it is expected that the environment has less influence on the verbal-short term memory task than the receptive vocabulary task. As we included age in our statistical model, we chose to use raw scores as well in order to minimize the comparison with French monolinguals.

[3]This sub-task aims to assess narrow phonological awareness.

[4]# designates a word-boundary.

Table 1. Structural properties of the analyzed nonwords.

Branching Onset (n = 31)	Final Lateral (n = 7)	Internal lateral Coda (n = 4)	Obstruent coda (n = 12)	#sC clusters (n = 9)
/pl/ - e.g., [pliks]	/l/ - [plal]	/l/ - [pilfu]	/k/ - [pliks]; [piks]	[skapufi]
[fupli]	/l/ - [pakifal]	/l/ - [filpa]	/p/ - [klups]; [fips]	[spu]
/kl/ -e.g., [klisp]	/l/ - [klil]	/l/ - [kupalfi]	/s/ - [pafuski]; [kuspa]	[skafu]
[paklu]	/l/ - [fikapul]	/l/ - [kufalpi]		[skla]
[piklafu]	/l/ - [fal]			[spli]
/fl/ - e.g., [flis]	/l/ - [fapul]			[spakifu]
[flukif]	/l/ - [kufal]			[spiku]
[kupifla]				[skap]
				[spaf]

Analyses

Focus on the correct repetition of complex syllables

Before modelling the correct repetition of whole nonwords, we first investigated the correct repetition of codas, final consonants, #sC clusters, and branching onsets regardless of the performance for other parts of the nonwords. Examples of nonwords containing those structures are given in Table 1.

We compared the medians of the five (non-normally distributed) distributions of by-subject performance – percentages of correct repetition of branching onsets, obstruent codas, coda /l/, final /l/, and #sC clusters) – with a Kruskal-Wallis test, post-hoc Dunn's tests, and Glass rank biserial correlation coefficients to measure effect size. We then compared the variance of the distributions with Fligner-Killeen tests and a Holm-Bonferroni *p*-value correction for multiple-pair testing.

Regression analysis of the subjects' performance

To assess our hypotheses, we considered a logistic regression model with mixed effects. The dependent variable is repetition, with two values: 0 for failure to repeat the nonword, and 1 for success. We included several independent variables, which were chosen according to the literature but also the preliminary analysis, exposed above, of the repetition of complex syllables:

- branching_onset - the number of branching onsets in the nonword (0, 1 or 2) (categorical)
- occurrence_l, which corresponds to whether we have an /l/ in final or coda position in the nonword (*coda* for /l/ in internal coda position, *final* for /l/ in final position, and *other* for /l/ neither in coda nor final position) (categorical)
- V – the number of vowels of the nonword (1, 2 or 3) (categorical)
- rec_voc – the size of a child's French receptive vocabulary (numerical)
- verbal_stm – the size of a child's verbal short-term memory (numerical)
- LoE – the length of a child's exposure to French (in days) (numerical)
- age, which is a child's age (in months) (numerical)
- L1_syll_complexity – the complexity of the syllables in the child's L1, as described in the Lapsyd database (Maddieson et al., 2013) (for bilingual children, the higher complexity

among the two L1 languages was considered) (numerical score between 0 and 8, see Table S2 in the supplementary materials for the values for the different languages and more details about how these values were computed)
- phono_awareness – the child's phonological awareness (z-scored performance for different tasks depending on the child's age) (numerical)

In addition to the main effects for these predictors, we hypothesized a number of meaningful interactions: branching_onset x LoE, occurrence_l x LoE, rec_voc x LoE, age x verbal_stm, age x rec_voc, and verbal_stm x rec_voc. These interactions corresponded to the following questions: i) Does the influence of syllable structures and of the size of the receptive vocabulary depend on LoE? ii) Does age modulate the influence of the verbal short-term memory and of the size of the receptive vocabulary? iii) Does the impact of the verbal short-term memory on NWR depends on the size of the receptive vocabulary?

Besides the previous fixed effects, we also considered random effects. Starting from a model with random intercepts only – to account for the groupings of observations according to participants (subject), their L1, and the stimuli to repeat (nonword) –, we adopted a step-up approach to find the best random effects structures. This led us to include by-subject random slopes for occurrence_l.[5]

Our approach is confirmatory, i.e., we wish to assess our hypotheses and only these hypotheses, which we posit are meaningful. We did not assess other possible interactions, nor conducted any model selection for the fixed effects.

All our hypotheses are oriented, as we felt it was reasonable, given the literature, to disregard alternative hypotheses. For

---

[5]The model can be described as follows, adopting the usual syntax in lme4: repetition ~ occurrence_l + branching_onset + V + LoE + age_months + rec_voc + verbal_stm + L1_syll_complexity + phono_awareness + occurrence_l:LoE + branching_onset:LoE + rec_voc:LoE + age_months:verbal_stm + age_months:rec_voc + rec_voc:verbal_stm + (0 + occurrence_l | subject) + (1 | L1) + (1 | nonword) [Family: binomial (logit)]. All continuous predictors were scaled in the model to ease convergence. We verified the assumptions of the model. We found high multicollinearity for LoE and occurrence_l:LoE but concluded false negative results for these predictors were unlikely. We worked with R (R Development Core Team, 2021) and most notably the packages lme4 (Bates et al., 2015) (glmer() function), lmerTest (Kuznetsova et al., 2017), emmeans (Lenth, 2022), sjPlot (Lüdecke, 2021) and DHARMa (Hartig, 2022).
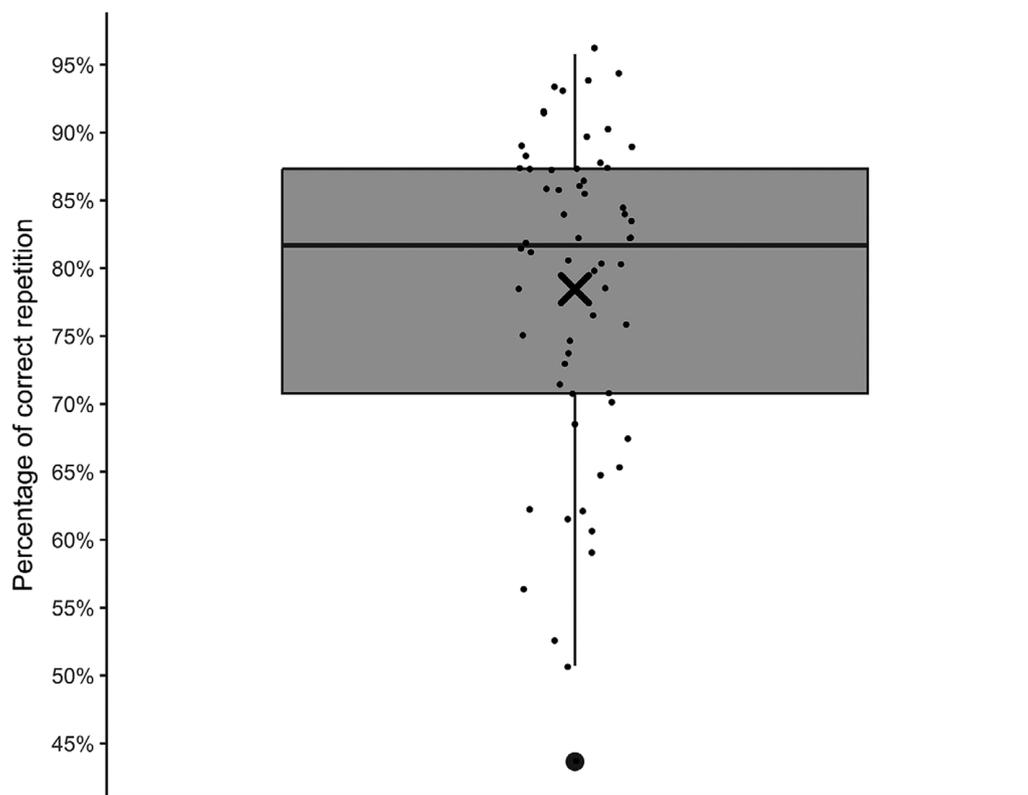
Figure 1. Inter-subject variability in performance for the nonword repetition task. Boxplot of the distribution of the participants' percentages of correct repetition for the nonword repetition task (LITMUS-QU-NWR-FR). The cross represents the average value of the distribution, the line in the box the median. The thicker dot represents an outlier – a performance value lower than 1.5 times the inter-quartile range from the value of the first quartile.

instance, while it suggests LoE may have either a positive impact or no impact on the children's performance, no study suggests a negative impact, and our corresponding (oriented) hypothesis is therefore that the more exposure to French, the easier to repeat the nonwords. We thus relied on unilateral rather than bilateral tests to assess the main effects of our statistical model and increase statistical power.

We assessed effects based on estimated marginal means (for categorical predictors) and estimated marginal trends (for numerical predictors). For multiple tests with categorical predictors, we relied on a multivariate *t* distribution with the same covariance structure as the estimates to adjust the *p*-values as exactly as possible.

All our computations are available with step-by-step explanations on Github at https://github.com/keruiduo/ChildL2AcquisitionFrench.

Results

We first report on bilingual children's overall performance on the LITMUS-QU-NWR-FR task. We then focus on performance at specific syllable structures to identify the areas of greater difficulties for the children in our study. Finally, we report on which factors influence children's performance.

Overall performance on the LITMUS-QU-NWR task

Figure 1 reports the percentage of correct repetition of the nonwords in the LITMUS-QU-NWR-FR task for our target group.

60% of the children (37/62) performed above 80% of target repetition, reaching the performance of typically developing French monolingual children. Crucially, they fell above the cut-off score separating typically developing children from children with developmental language disorder with good levels of specificity (79%) and sensitivity (81%) (Almeida et al., 2017). If we pay attention to the deciles, we can see that only 20% of the children perform below 69% of target repetition. This result is particularly impressive if we contrast it with children's performance on the receptive vocabulary task (see Figure S3 in the supplementary materials for the distribution of z-scores): none of the children reached French monolingual norms and the z-scores are very low, probably because the task is standardized for French monolinguals with much more exposure to French than our children ($MIN = -30.82$, $MAX = -3.56$, $M = -13.79$, $SD = 6.13$) (typically, the diagnosis of language impairment occurs when children perform lower than, depending on the tasks, 1.25 to 2 SD at standardized tasks. In other words, a score lower than -2 SD is very low).

Despite a high overall performance, a lot of individual variation is observed for the children's performance on the NWR task: even if the mean is high (78.46%) the percentage of correct repetition varies from 44% to 96%.

Focus on complex syllables

To better understand children's difficulties on the NWR task, we now focus on their production of complex syllable structures: branching onsets, lateral and obstruent codas, #sC clusters (i.e., word-initially) and word-final consonants. Figure 2 reports children's performance on these syllable constituents, independently of their performance on the nonword containing them, as it
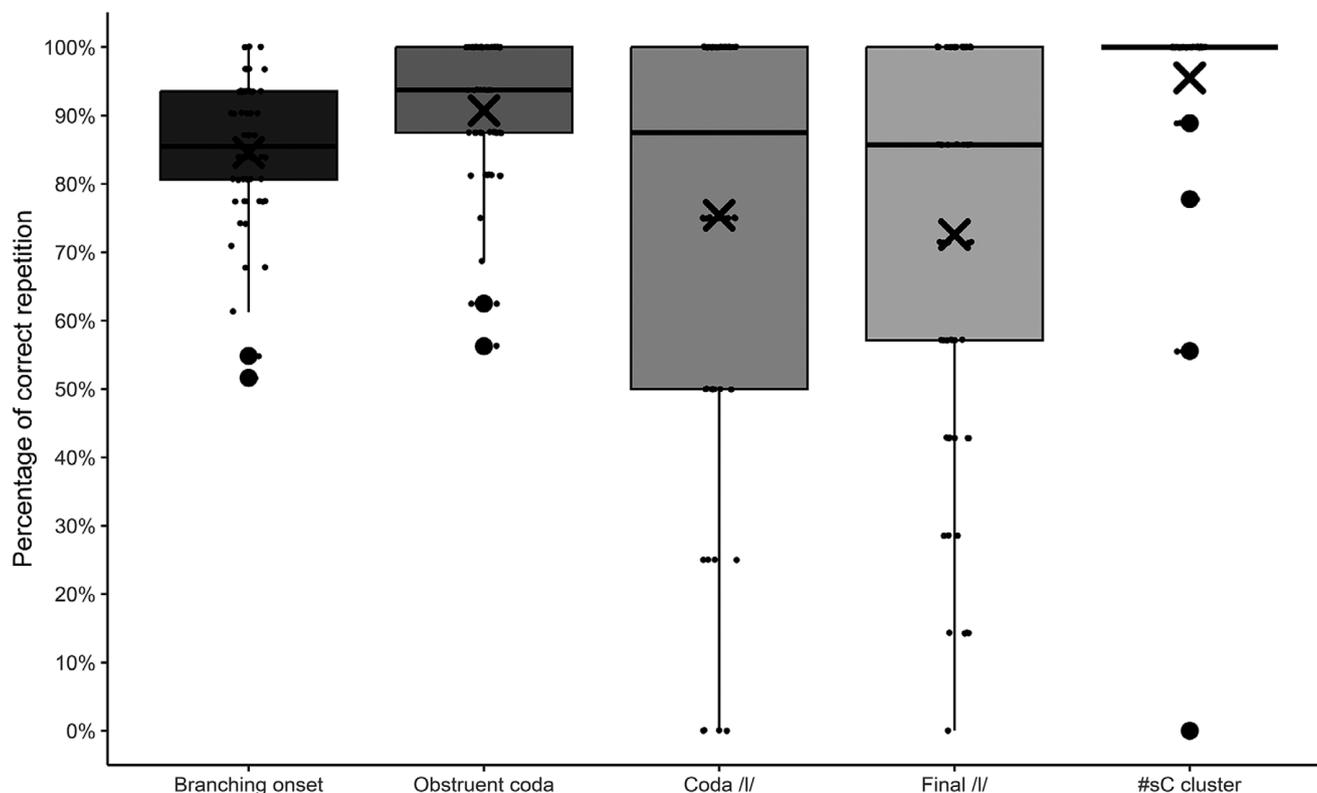
Figure 2. Impact of different syllable structures on correct repetition. Boxplots of the by-subject percentages of correct repetitions for five types of structures: branching onsets, obstruent codas, /l/ in internal coda position, /l/ in final position and #sC clusters. A cross represents the average value of a distribution, the line in a box the median. The thicker dots represent outliers, i.e., values further than 1.5 times the inter-quartile range from either the first or third quartile.

Table 2. Values and statistical assessment of the pair differences for the medians and variances of the 5 target syllabic structures

	median		variance
Syllable type			
Branching onset	0.855		0.012
Obstruent coda	0.938		0.014
Coda /l/	0.875		0.095
Final/l/	0.857		0.082
#sC cluster	1.000		0.023

Pair difference	z ratio	p-value	r_{rb}	Lower CI	Upper CI	χ^2	p-value
Branching onset - Obstruent coda	−3.460	**.004**	−.438	−.618	−.236	0.1	.746
Branching onset – Coda /l/	−1.411	.475	−.037	−.286	.187	45.4	**< .001**
Branching onset – Final /l/	−0.081	.935	.102	−.118	.325	25.0	**< .001**
Branching onset - #sC cluster	−6.475	**< .001**	−.762	−.896	−.614	32.5	**< .001**
Obstruent coda – Coda /l/	2.050	.162	.182	−.025	.372	45.8	**< .001**
Obstruent coda – Final /l/	3.379	**.004**	.346	.150	.528	30.5	**< .001**
Obstruent coda – #sC cluster	−3.015	**.013**	−.383	−.532	−.222	24.1	**< .001**
Coda /l/ - Final /l/	1.329	.367	.110	−.086	.313	3.46	.126
Coda /l/ - #sC cluster	−5.065	**< .001**	−.399	−.553	−.252	53.7	**< .001**
Final /l/ - #sC cluster	−6.394	**< .001**	−.536	−.666	−.389	39.9	**< .001**

Notes. (i) Both median and variance (of the percentages of correct repetition) are considered. (ii) rrb stands for Rank-biserial correlation. (iii) For pair differences, p-values are adjusted with the Holm-Bonferroni correction for both the medians and the variances. (iv) Significant p-values (< .05) are indicated in bold.

Table 3. Assessment of the hypothesized interactions between predictors in the logistic regression model.

Variable	Contrast	Estimate (β)	SE	asymp LCL	asymp UCL	z ratio	*p*-value
verbal_stm:rec_voc		7.87e–03	3.29e–02	−5.65e–02	7.23e–02	0.240	.811
rec_voc:age		8.50e–04	1.56e–03	−2.21e–03	3.91e–03	0.544	.586
verbal_stm:age		−3.22e–03	1.21e–02	−2.69e–02	2.05e–02	−0.266	.790
rec_voc:LoE		−2.90e–04	2.11e–04	−7.04e–04	1.23e–04	−1.377	.168
occurrence_l:LoE	coda - final	−2.33e–03	4.39e–03	−1.25e–03	7.85e–03	−0.531	.850*
occurrence_l:LoE	coda - other	1.19e–03	2.85e–03	−5.41e–03	7.80e–03	0.419	.904*
occurrence_l:LoE	final - other	3.52e–03	3.48e–03	−4.56e–03	1.16e–03	1.011	.557*
branching_onset:LoE	0–1	1.85e–03	1.16e–03	−8.09e–04	4.52e–03	1.603	.230*
branching_onset:LoE	0–2	1.57e–04	2.66e–03	−5.96e–03	6.28e–03	0.059	.998*
branching_onset:LoE	1–2	−1.70e–03	2.64e–03	−7.77e–03	4.38e–03	−0.642	.787*

Notes. (i) Estimated marginal trends, standard errors, control limits (LCL/UCL), z ratio and p-values obtained with the function emtrends() ('emmeans' package) applied to our glmer model. (ii) *: multivariate t-distribution method for confidence-level and p-value adjustment. (iii) two-tailed p-values.

Table 4. Assessment of the categorical main effects in the logistic regression model.

Variable	Contrast	Estimate (β)	SE	asymp LCL	asymp UCL	z ratio	*p*-value
occurrence_l	coda - final	−0.489	0.526	-Inf	0.604	−0.930	.396
occurrence_l	coda - other	−1.244	0.398	-Inf	−0.417	−3.129	**.002**
occurrence_l	final - other	−0.756	0.373	-Inf	0.020	−2.026	.056
branching_onset	0–1	0.847	0.184	0.477	Inf	4.614	**< .001**
branching_onset	0–2	1.875	0.499	0.869	Inf	3.760	**< .001**
branching_onset	1–2	1.028	0.506	0.008	Inf	2.033	**.048**
V	1–2	0.077	0.212	−0.365	Inf	0.361	.667
V	1–3	0.576	0.217	0.126	Inf	2.660	**.011**
V	2–3	0.500	0.215	0.053	Inf	2.324	**.027**

Notes. (i) Differences between estimated marginal means of the contrasted levels, standard errors, control limits (LCL/UCL), z ratio and p-values obtained with the function emmeans() ('emmeans' package) applied to our glmer model. (ii) multivariate t-distribution method for confidence-level and p-value adjustment. (iii) one-tailed p-values according to oriented hypotheses. (iv) Significant p-values (< .05) are indicated in bold.

Table 5. Assessment of the continuous main effects.

Variable	Estimate (β)	SE	asymp LCL	asymp UCL	z ratio	*p*-value
rec_voc	0.055	0.021	0.021	Inf	2.672	**.004**
age	7.16e–03	0.012	−0.012	Inf	0.619	.268
verbal_stm	−8.26e–03	0.123	−0.211	Inf	−0.067	.527
LoE	1.58e–03	1.99e–03	−1.70e–03	Inf	0.791	.214
L1_syll_complexity	0.159	0.134	−0.062	Inf	1.184	.118
phono_awareness	0.125	0.068	0.013	Inf	1.837	**.033**

Notes. (i) Estimated marginal trends, standard errors, control limits (LCL/UCL), z ratio and p-values obtained with the function emtrends() ('emmeans' package) applied to our glmer model. (ii) One-tailed p-values according to oriented hypotheses. (iii) Significant p-values (< .05) are indicated in bold.

can happen that a nonword is not correctly produced but these particular constituents are.

Average percentages of correct repetition are equal to 84.6%, 90.7%, 75.4%, 72.6%, and 95.5% for branching onsets, obstruent codas, coda /l/, final /l/, and #sC clusters, respectively. These values are, however, little informative given the strong heterogeneity among children and the ceilinged distributions: scores range for instance from 0 to 100% for coda /l/ and final /l/, and from around 50% to 100% for branching onsets. In terms of errors, we can group the structures for which the segments are preserved and the

structures for which segments are affected. In the case of #sC clusters, final /l/ and coda /l/, the segments are preserved. For #sC clusters, the main repair strategy is vowel insertion initially (59%;17/29) (e. g. /skapufi/ → [øskapofi]) and for final /l/, vowel insertion word-finally (97%;159/164), (e. g. /kufal/ → [kufalə]). In the case of coda /l/, the main strategy is metathesis (59%;34/58)), (e. g. /kufalpi/ → [kufapli]). In the case of obstruent codas, the main repair strategy is substitution of the obstruent (48%;12/25)) (e. g. /kusp/ → [kufp]) and for branching onsets, deletion of the lateral (33%;100/301), (e. g. /plifu/ → [pifu]). In brief, errors affect
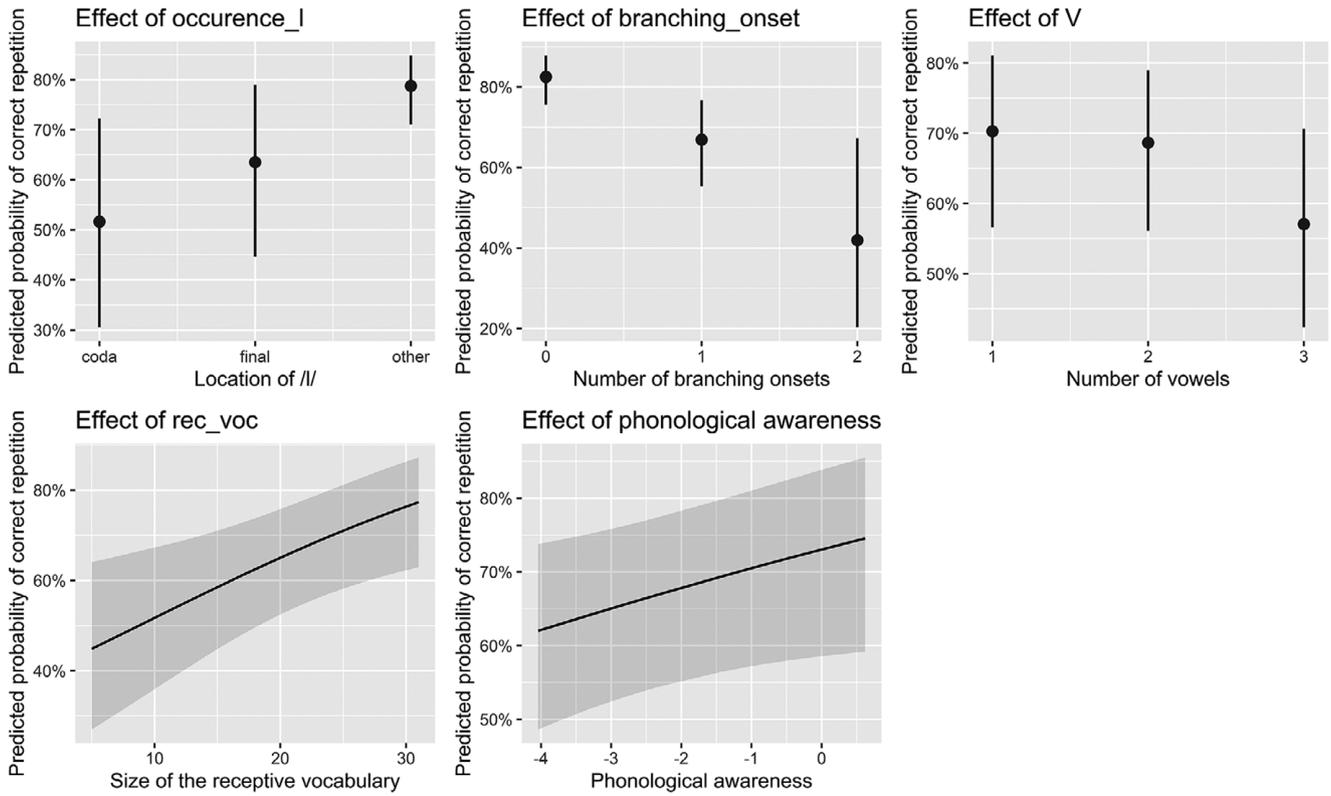
Figure 3. Impact of the significant predictors on correct nonword repetition. Combined plots of the impact on the percentage of correct repetition of 5 predictors with either statistical significance or a statistical tendency: occurrence_l (categorical), branching_onset (categorical), V (categorical), rec_voc (continuous), age (continuous) and phono_awareness (continuous). Estimated marginal means and marginal trends derived from the mixed-effects logistic regression model are shown along with their confidence intervals.
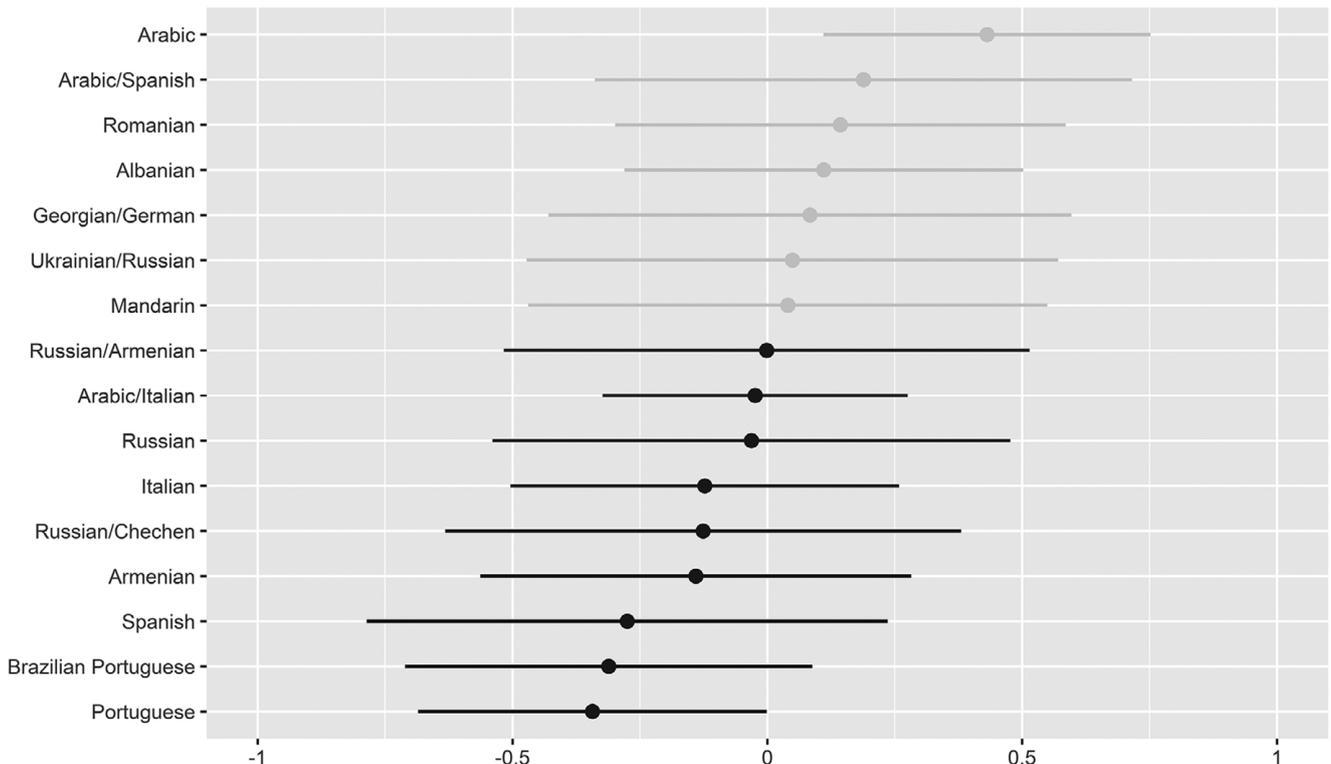

Figure 4: Impact of L1 on the percentage of correct nonword repetition. Plot of the distribution of the estimated values of the random intercept for L1 (16 levels), along with their confidence intervals.

segments for branching onsets and obstruent codas, contrary to #sC clusters, coda /l/ and final /l/, where segments are preserved. Statistical inferences can be made to go further. Table 2 reports both median and variance for the 5 syllable structures, as well as the possible pairs among them with assessment of significant differences.

Medians for the five syllable structures confirm observations with means: #sC clusters and obstruent codas lead to better performances than branching onsets, coda /l/ and final /l/, which have close median percentages of correct repetition. Further, when it comes to variance, two groups can be distinguished: coda /l/ and final /l/ on the one hand, with higher variances (Mdn = 0.095 and 0.082, respectively), and on the other hand #sC clusters, obstruent codas and branching onsets, with lower variances (Mdn = 0.023, 0.014 and 0.012, respectively).

A Kruskal-Wallis test reveals significant differences between the medians of the five distributions ($\chi^2 = 59.56$, $p < .001$). Post-hoc Dunn's tests and rank biserial correlations show significant differences between #sC clusters and all four other structures: branching onsets, $z = -6.47$, $p < .001$, $r_{rb} = -.762$ (large), coda /l/, $z = -5.06$, $p = .004$, $r_{rb} = -.399$ (medium), final /l/, $z = -6.39$, $p < .001$, $r_{rb} = .536$ (large), and obstruent codas, $z = -3.01$, $p = .004$, $r_{rb} = -.383$ (medium). They also show a significant difference between obstruent codas and branching onsets, $z = -3.46$, $p = .004$, $r_{rb} = -.438$ (medium), as well as between obstruent codas and final /l/, $z = 3.38$, $p = .004$, $r_{rb} = .346$ (medium). We do not observe, however, significant differences between branching onset, coda /l/ and final /l/. There is also no significant difference between coda /l/ and obstruent codas, although there would be one without adjustment of the p-values.

In terms of variance, Fligner-Killeen tests suggest significant differences between all pairs of structures except between coda /l/ and final /l/, and between obstruent codas and branching onsets.

Overall, one can argue that obstruent codas and #sC clusters are not problematic for the children (higher medians and lower variances), while coda /l/ and final /l/ are more challenging for at least some of the children (lower medians and higher variances). Branching onsets fall in between with a lower median and a lower variance, but one may arguably group them with coda /l/ and final /l/ given their medians are not significantly different. These results explain why we focus on these three structures in our regression analysis and leave aside #sC clusters and obstruent codas.

Determinants of NWR performance

The mixed-effects logistic regression model designed to assess the influence of various predictors on NWR performance contained six interactions, but we found that none of them was significant (see Table 3). We therefore focused on the main effects.

Our different hypotheses were not all confirmed but we found several significant effects (see Table 4, Table 5, and Figure 3). There is first an inhibitory effect of the occurrence of /l/, with nonwords significantly harder to repeat when /l/ appears in internal coda position than when there is no /l/ in coda or final position (condition 'other') ($\beta = -1.244$, $z = -3.129$, adj. $p = .002$). One clear tendency ($p < .1$) was additionally observed, with nonwords harder to repeat when /l/ appears in final position than when there is no /l/ in coda or final position ($\beta = -0.756$, $z = -2.026$, adj. $p = .056$). Second, there is a significant inhibitory effect of branching_onset: i) nonwords without branching onset are easier than nonwords with one branching onset ($\beta = 0.847$, $z = 4.614$, adj. $p < .001$), ii) nonwords without branching onset are easier than

nonwords with two branching onsets ($\beta = 1.875$, $z = 3.760$, adj. $p < .001$), and iii) nonwords with one branching onset are easier than nonwords with two branching onsets ($\beta = 1.028$, $z = 2.033$, adj. $p = .048$) (5). Third, we observe an inhibitory effect of V: nonwords with one vowel are easier than nonwords with three ($\beta = 0.576$, $z = 2.660$, adj. $p = .011$), while nonwords with two vowels are easier than nonwords with three ($\beta = 0.500$, $z = 2.324$, adj. $p = .027$). Finally, there are two facilitatory effects or rec_voc and phono_awareness: the larger a child's receptive vocabulary, the easier to repeat the nonwords ($\beta = 0.055$, $z = 2.672$, $p = .004$), and the more developed their phonological awareness, the easier the task is too ($\beta = 0.125$, $z = 1.837$, $p = .033$).

The random intercepts for L1 offer a glimpse into the possible impact of different L1s on the repetition task (see Figure 4). While one may hypothesize that being bilingual before learning French helps with the repetition task, the distribution of the intercepts is not suggestive of such an effect: for instance, among the five L1 configurations with the strongest intercepts, i.e., the strongest facilitatory effects on correct repetition, only two correspond to bilingual children (Arabic/Spanish and Georgian/German).

Discussion

The main goal of this article was to evaluate eL2 phonological performance on the LITMUS-QU-NWR-FR task during children's first year of exposure to French. Our results show that the majority of the children (37 out of 62 children, i.e., 60%) reach 80% of correct repetition at the task. Despite this, a lot of variation was found, especially for the production of complex syllables, namely branching onsets, coda /l/ and final /l/. The presence of these three syllable structures in the nonword impacted NWR performance, as did the increase in the number of vowels. A greater phonological awareness score and a greater L2 receptive vocabulary score impacted NWR positively. Verbal short-term memory, independently assessed by a forward digit span task, as well as L1, age and LoE, did not have an impact on NWR.

Overall performance

As 87% of eL2 English French learners reached 80% of accuracy at the LITMUS-QU-NWR-FR task within several months of exposure (Scheidnes, 2020), we expected our children to do the same. This hypothesis was verified, since 60% reached 80% of target repetition, which is the cut-off score previously established for this task that allows to separate children with and without developmental language disorder, independently of bilingualism (only 21% of the children scored below 80% in Almeida et al., 2017). The main challenge of assessing bilingual children's language development is that they can exhibit a lower performance on a particular language even if they do not have a language impairment, but only because of little exposure to the L2 (Armon-Lotem et al., 2015). Considering that only 20% of our children scored below 69%, one possibility would be to lower the cut-off for this specific population and set it to 70%. In order to do that, we would need a group of eL2 with developmental language disorder and test the sensitivity of the task with different cut-offs, which goes well beyond the scope of the present paper.[6] The fact that, in our study, 80% of the children with

---

[6]As it is difficult to diagnose bilingual children with developmental language disorder, this would imply to focus on children with an established diagnosis in their L1, before exposure to their L2.

very little exposure to the L2 with different L1 backgrounds reach 69% of target repetition, i.e., values above or close to the cut-off previously established for monolinguals and bilinguals with no language impairment is, thus, a good indicator that the LITMUS-QU-NWR-FR task can contribute to assess language impairment in this specific population.

Performance on complex syllable structures

As L2 children have been reported to exhibit difficulties at specific structures even when they have an overall good L2 phonological system (Scheidnes, 2020), we expected the performance on complex syllables to be below 80%. This hypothesis was partially verified since only some complex syllables were problematic: obstruent codas and #sC clusters reached very high levels of accuracy while branching onsets, coda /l/ and final /l/ were more a source of difficulty for them. This result expands previous research on NWR tasks, as most studies revealed difficulties with clusters in general (Scheidnes, 2020) or codas, without separating liquid and obstruent codas or medial and final codas (e.g., Duncan & Paradis, 2016). Here we showed that the segmental quality of the coda impacts children's performance since they are better at obstruent codas than at lateral codas, exhibiting a pattern of preference for less sonorous segments in word-medial coda position attested in L1 acquisition for French (Kehoe, 2021) English (Kehoe & Stoel-Gammon, 2001), Dutch (Fikkert, 1994), and European Portuguese (Freitas, 1997). This result adds to previous evidence showing difficulties with /l/ in coda position in bilingual children with and without developmental language disorder (Almeida et al., 2019; Ferré et al., 2015) and eL2 children (Scheidnes, 2020). The fact that this result was attested in other studies contributes to its robustness across different populations.

Impact of internal and external variables on NWR repetition

Our children exhibited a lot of variation at the task, contrary to what had been reported with previous studies using the LITMUS-QU-NWR task (Grimm, 2022, Abed Ibrahim et al., 2020; a.o.). In the next subsections, we discuss the factors that may explain this variation.

Phonological complexity

Our statistical analyses highlighted that the presence of /l/ in branching onset, word-medial coda and word-final position impacted NWR negatively. When this happened, the error could affect the lateral, leading to its deletion or substitution, but in some cases the lateral was preserved but produced in another place of the nonword or followed by vowel epenthesis. In yet some other instances, the error did not affect the target structure but implied a non-correct repetition of the whole item (e. g. /fupli/ → [flupli]). While difficulties with /l/ in coda had been previously established for this task (Ferré et al., 2015), difficulties with final /l/ had not previously been reported, to our knowledge. Moreover, while difficulties with branching onsets had been described for NWR tasks (Almeida et al., 2019; Duncan & Paradis, 2016), here we could find that nonwords with two branching onsets were more difficult than nonwords with only one branching onset, following previous evidence that adding complexity to items makes them more difficult to repeat (Gallon et al., 2007). This result must be interpreted with caution, though, as only 2 items have two branching onsets and 29 have only one. This result should be replicated with a task containing the same number of items with 1 and 2 branching

onsets. Another important finding is that branching onsets with a lateral, coda /l/ and final /l/ are more difficult than #sC clusters and obstruent codas. There are very few data on the acquisition of word-initial #sC clusters in French. The data from dos Santos (2007) suggest that #sC clusters may be acquired after branching onsets but this result is based on a longitudinal case-study. Note, however, that an earlier acquisition of #sC clusters in comparison with branching onsets has been reported for European Portuguese (Freitas, 1997). Further studies should focus on the development of #sC clusters in monolingual French children in order to identify an age/pattern of acquisition for this structure.

We expected that a greater phonological awareness would positively impact NWR, as previous literature had shown that L1 phonological awareness positively impacted L2 sound structure acquisition (Engel de Abreu & Gathercole, 2012). This hypothesis was verified in our data. This finding is new since it shows that L2 phonological awareness, and not only L1 phonological awareness, promotes eL2 NWR. This finding can have important clinical implications: while speech and language therapists often use phonological awareness to improve phonological acquisition, it is unlikely that they are able to do so with L1 phonological awareness in the case of immigrated children. Our results strongly suggest that working with L2 phonological awareness may promote L2 phonological acquisition, even in children with little exposure to the L2. One limitation to keep in mind is, however, that the sub-task we relied on with our participants measured narrow phonological awareness rather than a broader capacity.

L2 receptive vocabulary

We expected that a large French receptive vocabulary would promote NWR and this hypothesis was indeed verified. Duncan and Paradis (2016) is the only study who found an effect of eL2 vocabulary size on NWR performance. As their task relies primarily on verbal short-term memory, such an influence is expected, since verbal short-term memory and vocabulary knowledge are reportedly linked during acquisition (Gathercole et al., 1999). The LITMUS_QU-NWR task has the particularity of being constructed out of segments in accordance with French phonotactic restrictions, and not via existing words or morphemes, which is often the case with other NWR tasks. Our result thus reinforces that receptive vocabulary positively influences the acquisition of eL2 phonology, even in a task that relies less on lexical knowledge than other NWR tasks. To our knowledge, our study is the first to report such a link between vocabulary and the LITMUS-QU-NWR task.

Age

We expected that older children would be better at the NWR task. We did not find such an effect, which goes against previous findings (e. g. Duncan & Paradis, 2016). This might be explained by the fact that LITMUS-QU-NWR tries to minimize the impact of verbal short-term memory, by including NWR up to three syllables maximum. Duncan and Paradis (2016) stated that the influence of age is due to the maturation of verbal short-term memory, and they used a NWR task designed to assess verbal short-term memory, using long nonwords. Consequently, an influence of age on children's NWR task was totally expected. As we saw, LITMUS-QU-NWR-FR relies much less on verbal short-term memory, as items are shorter. The impact of variables related to verbal short-term memory seems thus to be minimized for the LITMUS-QU-NWR task in comparison with other NWR tasks.

Verbal short-term memory

Even if the LITMUS-QU-NWR task minimizes the effects of verbal short-term memory, as shown in the previous section, such effects cannot be entirely ruled out, as dos Santos & Ferré (2018) pointed out. We thus expected an effect of verbal short-term memory on the task and investigated it in two complementary ways: within the task itself, we expected that items with 3 vowels made repetition more challenging than items with 1 or 2 vowels. We also independently assessed children's verbal short-term memory and we expected that the larger it is, the better NWR. Only the former hypothesis was verified in our data. Dos Santos and Ferré (2018) found an effect of the number of vowels at this particular task: children's performance tended to decrease between dissyllabic and trisyllabic non-words (but there was no change between monosyllabic and dissyllabic nonwords). We found the same result even if the population observed was different. This decrease seems to be related to verbal short-term memory effects, as such effects typically start with items with three syllables. If we accept that verbal short-term memory has an impact on children's performance on the LITMUS-QU-NWR-FR task, by the influence of the number of vowels, it is surprising that the forward digit span task that we used to specifically assess children's verbal short-term memory did not impact children's accuracy at the NWR task. A possible explanation for that may be the fact that children were asked to do the forward digit span task in French, which is their L2. Even if the knowledge of the digits in French did not apparently seem to be a problem for the children, it is still possible that repeating a digit in a language for which our phonological representations are not fully stable is a difficult task. In other words, it is possible that the children's results at the forward digit span task do not reflect their verbal short-term memory skills, and we should then evaluate children's verbal short-term memory capacities through tasks in their L1.

L1 syllabic complexity

We expected that the more complex the syllables in the child's L1, the easier correct repetition. This hypothesis was ruled out in our statistical analysis. This finding goes against previous literature showing that children with a L1 with high complexity on clusters performed better than children having a L1 with low complex syllable properties (Duncan & Paradis, 2016). This result was further replicated in Kehoe & Havy (2019) for simultaneous bilinguals. Our result is thus surprising, especially considering that Duncan and Paradis (2016) state that the influence of L1 was most pronounced for children in the early stages of L2 exposure, like ours. On the other hand, Chilla et al. (2021) did not find any influence of L1 on the LITMUS-QU-NWR task for simultaneous and successive learners with at least one year of exposure to the L2. Our result is thus in line with previous research on the LITMUS-QU-NWR task and expands it showing no L1 effect even with 18 different languages (and not only three). Moreover, it shows no influence of L1 in the very first stages of L2 acquisition, a setting in which L1 influence is particularly expected. This result highlights the robustness of the LITMUS-QU-NWR task for use with a wide variety of bilingual children. Note, however, that to explore the impact of L1 influence, one should compare the properties of each specific structure in each language, a comparison impossible to conduct with so many different L1s. Instead, we used a global measure of syllabic complexity. Future results may be different from ours if they rely on a more detailed comparison between languages.

LoE

We expected that a longer exposure to French would promote NWR, following results from Duncan and Paradis (2016). This hypothesis was not confirmed in our data. It is possible that LoE did not impact our results because we assembled quite a homogenous group of children with respect to that measure: our participants vary in their LoE in months (between 1 and 11), and not in years, as was the case for the children evaluated by Duncan and Paradis (2016). Scheidnes (2020) did not find an effect of LoE on eL2 French phonological performance using a shorter version of the task we used, focusing also on children with low variation with respect to that measure (only a few months), like we did. The comparison of this last study with our own is particularly interesting because the eL2 children in Scheidnes (2020) are similar to ours in terms of L2 input quantity (as LoE was also very low) but not in terms of quality: the former were exposed to French through their teacher only, so that they had no varied input. On the contrary, our participants were exposed to French through a wide variety of people: their different French teachers and their schoolmates. Crucially, they had started living in a French-speaking environment. Despite this difference in quality of exposure, the absence of a LoE effect is consistent. This finding is quite robust for the LITMUS-QU-NWR tasks: these tasks are particularly poorly impacted by language experience, so it is possible that there will not be any effect of LoE even in children with longer exposure, contrary to the NWR task used by Duncan & Paradis (2016).

Limitations

This study has several limitations, the most important one being, in our opinion, the fact that we did not assess children's L1 phonological competence. This was not possible given the fact that children came from diverse L1 backgrounds. Nevertheless, given that children need so little exposure to the L2 to perform well at the NWR task, the next step would be to see if their L1 phonological representations play a role in their L2 phonological performance. Answering this question is primordial for theoretical purposes. Future research should also ideally consider socioeconomic status as a potential source for variation.

As the LITMUS-QU-NWR task assesses mainly syllable structure, it does not provide a complete picture of the children's L2 phonological competence. It does not evaluate all their phonetic inventory, as only 3 vowels and 5 consonants are represented in the task. Further, the overall accuracy measure does not give information about the fine-grained aspects of phonological competence, such as Voice Onset Time. In order to evaluate children's segmental phonology, another task should be used. Moreover, the LITMUS-QU-NWR task allowed us to analyze the production of complex syllable structures with /l/ but this analysis is based on unequal numbers of items across condition. Ideally, the same number of items for different conditions should be used.

Finally, we wish to stress that, even though the LITMUS-QU-NWR-FR task seems useful in assessing L2 phonology during the first steps of its acquisition, direct assessment of language disorders cannot be based on phonology alone: it requires the evaluation of other language domains, in combination with parental reports about early development (Boerma & Blom, 2017).

Notwithstanding, using this task with newly immigrated children can be useful as a first assessment when assessing other language domains seems not feasible. To further evaluate its accuracy in identifying language impairment in eL2 children, we would need to have a longitudinal design and follow the evolution of these children's language development.

Conclusion

Our study is, to our knowledge, the first to focus on NWR accuracy during the first year of exposure to an L2 (French) in a population of newly immigrated children with different L1 backgrounds. By doing so, we wanted to test the possibility of assessing bilingual children's phonological performance in the L2 very early, contributing empirical evidence to our current theoretical and clinical knowledge. Our results show that, contrary to vocabulary, phonological performance, measured in terms of syllable accuracy, can be assessed shortly after the onset of exposure to the L2. This is an important finding both theoretically, as it contributes to characterize the specificities of eL2 acquisition, namely the fact that L2 syllable structure develops fast, and clinically, as it strongly suggests that LITMUS-QU-NWR-FR can contribute to identify language disorders in eL2 children. Finally, our study reinforces the potential of a tool such as LITMUS-QU-NWR-FR to assess phonological development in diverse situations of bilingualism, as this task seems neutral to L1 and LoE.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/S136672892400083X.

Data availability statement. Supporting data are available at https://github.com/keruiduo/ChildL2AcquisitionFrench. The raw transcriptions will be made available in PhonBank (https://phon.talkbank.org/access/) within the next 3 months after publication of this article.

References

Abi-Aad, K., & Atallah, C. (2020). L'épreuve répétition de non-mots: LITMUS-NWR-LIBAN. In Zebib, R., Prévost, P., Tuller, L. & Henry, G. (Eds.), *Plurilinguisme et Troubles Spécifiques du Langage au Liban*, pp. 79–92. Beyrouth: Presses universitaires de l'Université Saint-Joseph.

Abed Ibrahim, L., Hamann, C., & Fekete, I. (2020). *Language Assessment of Bilingual Arabic-German Heritage and Refugee Children: Comparing Performance on LITMUS Repetition Tasks. In* Brown, M. M & A Kohut (eds.) Proceedings of the 44th Boston University Conference on Language Development, pp. 1–17. Somerville, MA: Cascadilla Press.

Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A.(2017). Identification of Bilingual Children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism* 7 (3–4), 331–358. https://doi.org/10.1075/lab.15019.alm

Almeida, L., Ferré, S., Barthez, M.-A., & dos Santos, C. (2019). What do monolingual and bilingual children with and without SLI produce when phonology is too complex? *First Language*, **39**(2), 158–176. https://doi.org/10.1177/0142723718805665

Armon-Lotem, S. & Grohmann, K. (2021). *Language Impairment in Multilingual Settings*. LITMUS Action across Europe. John Benjamins. https://doi.org/10.1075/tilar.29

Boersma, P. & Weenink, D. (2019). Praat doing phonetics by computer [Computer program].: Version 6.0.52, retrieved 20 Mai 2019 from http://www.praat.org/

Bates, D., Maechler, M., Bolker, B. M. & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boerma, T. & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders* **66**, 65–76.

Chevrie-Muller, C. & Plaza, M. (2001). N-EEL - Nouvelles épreuves pour l'examen du Langage. Retrieved from https://www.pearsonclinical.fr/n-eel-nouvelles-epreuves-pour-lexamen-du-langage

Chilla, S., Haman, C., Prévost, P., Abed-Ibrahim, L., Ferré, S., dos Santos, C., Zebib, R., & Tuller, L. (2021). The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German. In Armon-Lotem, S. & Grohmann, K. (Eds.). *Language Impairment in Multilingual Settings. LITMUS in action across Europe*, pp. 227–262. John Benjamins.

Cilibrasi, L., Stojanovik, V., Loucas, T., & Riddell, P. (2018). The role of noninitial clusters in the Children's Test of Nonword Repetition: Evidence from children with language impairment and typically developing children. *Dyslexia* **24**(4), 322–335. https://doi.org/10.1002/dys.1599

Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua* **95**, 5–26.

Duncan, T. S. & Paradis, J. (2016). English Language Learners' Nonword Repetition Performance: The Influence of Age, L2 Vocabulary Size, Length of L2 Exposure, and L1 Phonology. *Journal of Speech Language and Hearing Research* **59**(1), 39. https://doi.org/10.1044/2015_JSLHR-L-14-0020

Engel de Abreu, P. M. J. & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology* **104**(4), 974–986. https://doi.org/10.1037/a0028390

Ferré, S., dos Santos, C., & de Almeida, L. (2015). Potential phonological markers for SLI inbilingual children. In E. Grillo, & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development*, pp. 152–164. Somerville, MA: Cascadilla Press.

Fikkert, P. (1994). *On the Acquisition of Prosodic Structure*. Dordrecht: HIL.

Freitas, M. J. (1997). Aquisição da estrutura silábica do Português Europeu. Unpublished Ph.D Thesis. Univerity of Lisbon.

Gallon, N., Harris, J., & van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics & Phonetics* **21**, 435–455. https://doi.org/10.1080/02699200701299982

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (2007). The children's test of nonword repetition: A test of phonological working memory. *Memory* **2**(2), 103–127. https://doi.org/10.1080/09658219408258940

Gathercole, S. E., Service, E., Hitch, G. H., Adams, A.-M., & Martin, A. J. (1999). Phonological Short-term Memory and Vocabulary Development: Further Evidence on the Nature of the Relationship. *Applied Cognitive Psychology* **13**, 65–77.

Genesee, F. & Geva, E. (2006). Cross-Linguistic Relationships in Working Memory, Phonological Processes, and Oral Language. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 175–183). Lawrence Erlbaum Associates Publishers.

Grimm, A. (2022). The Use of the LITMUS Quasi-Universal Nonword Repetition Task to Identify DLD in Monolingual and Early Second Language Learners Aged 8 to 10. *Languages* **7**, 218. https://doi.org/10.3390/languages7030218

Grimm, A. & Domahs, U. (2023). The acquisition of consonant clusters and word stress by early second language learners of German: Evidence for cross-linguistic influence? *Linguistic Approaches to Bilingualism*, Online First. https://doi.org/10.1075/lab.21026.gri

Hamann, C. & Abed Ibrahim, L. (2017). Methods for Identifying Specific Language Impairment in Bilingual Populations in Germany. *Frontiers in Communication* **2**. https://doi.org/10.3389/fcomm.2017.00016

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. https://cran.r-project.org/package=DHARMa

Hedlund, G. & Rose, Y. (2020). *Phon 3.1* [Computer Software]. Retrieved from https://phon.ca

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The Relationship Between Phonological Awareness and Reading. *Language Speech and Hearing Services in Schools* **36**(4), 285–293. https://doi.org/10.1044/0161-1461(2005/029)

Hu, C.-F. (2003). Phonological Memory, Phonological Awareness, and Foreign Language Word Learning. *Language Learning* **53**, 429–462. https://doi.org/10.1111/1467-9922.00231

Kehoe, M. (2021). Coda consonant production in French-speaking children. *Clinical Linguistics & Phonetics* **35**(6), 509–533, https://doi.org/10.1080/02699206.2020.1795723

Kehoe, M. (2015). Cross-linguistic interaction: A retrospective and prospective view. In E. Babatsouli & D. Ingram (Eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*, pp. 141–167. Institute of Monolingual and Bilingual Speech.

Kehoe, M. & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. in *Journal of Child Language* **28**(2), 393–432. https://doi.org/10.1017/S030500090100469X

Kehoe, M. & Havy, M. (2019). Bilingual phonological acquisition: the influence of language-internal, language-external, and lexical factors. *Journal of Child Language* **46**(2), 292–333. https://doi.org/10.1017/S0305000918000478

Kuznetsova, A., Brockhoff, P.B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://cran.r-project.org/package=emmeans

Lüdecke, D. (2021). sjPlot: Data Visualization for Statistics in Social Science. https://cran.r-project.org/package=sjPlot

Maddieson, I., Flavier, S., Marsico, E., Coupé, C., & Pellegrino, F. (2013). LAPSyD: Lyon – Albuquerque Phonological Systems Database. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pp. 3022–3026.

Meisel, J. (2018). Early child second language acquisition: French gender in German children. *Bilingualism: Language and Cognition* **21**(4), 656–673. https://doi.org/10.1017/S1366728916000237

Meziane, R. S. & MacLeod, A. (2017). L'acquisition de la phonologie en français langue seconde : le profil phonologique d'enfants allophones en maternelle. *Canadian Journal of Applied Linguistics / Revue canadienne de linguistique appliquée* **20**(2), 1–17. https://doi.org/10.7202/1042673ar

Morrow, A., Goldstein, B. A., Gilhool, A., & Paradis, J. (2014). Phonological Skills in English Language Learners. *Language Speech and Hearing Services in Schools* **45**(1), 26–39. https://doi.org/10.1044/2013_LSHSS-13-0009

Paradis, J. (2016). An agenda for knowledge-oriented research on bilingualism in children with developmental disorders. *Journal of Communication Disorders* **63**, 79–84. https://doi.org/10.1016/j.jcomdis.2016.08.002

Piggott, G. (1999). At the right edge of words. *The Linguistic Review* **16**(2), 143–185. https://doi.org/10.1515/tlir.1999.16.2.143

R Development Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.r-project.org

Rattanasone, N. X. & Demuth, K. (2014). The acquisition of coda consonants by Mandarin early child L2 learners of English. *Bilingualism: Language and Cognition* **17**(3), 646–659. https://doi.org/10.1017/S1366728913000618

Rose, Y. (2000). *Headedness and Prosodic Licensing in the L1 Acquisition of Phonology*. PhD dissertation. Mc Gill University.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). Introducing Phon: A Software Solution for the Study of Phonological Acquisition. In *Proceedings of the Annual Boston University Conference on Language Development*, pp. 489–500.

dos Santos, C. (2007). Développement phonologique en français langue maternelle: Une étude de cas [Unpublished doctoral dissertation], *Université Lumière Lyon* 2. https://hal.archives-ouvertes.fr/tel-03752459

dos Santos, C. & Ferré, S. (2018). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition* **25**(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Scheidnes, M. (2020). Sentence repetition and non-word repetition in early total French immersion. *Applied Psycholinguistics* **41**, 107–131. https://doi.org/10.1017/S0142716419000420

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech Language and Hearing Research* **64**(5), 1–16. https://doi.org/10.1044/2021_JSLHR-20-00552

Schwob, S. & Skoruppa, K. (2022). Detecting Developmental Language Disorder in Monolingual and Bilingual Children: Comparison of Language-Specific and Crosslinguistic Nonword Repetition Tasks in French and Portuguese. *Journal of Speech, Language, and Hearing Research* **65**(3), 1159–1165. https://doi.org/10.1044/2021_JSLHR-21-00017

Selkirk, E. (1982). The syllable. In Hulst, HV and Smith, N (eds), *The structure of phonological representations: Part 2*, pp. 37–384. Foris.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language* **38**, 1–34.

Tessier, A.-M., Duncan, T. S., & Paradis, J. (2013). Developmental trends and L1 effects in early L2 learners' onset cluster production. *Bilingualism: Language and Cognition* **16**, 663–681.

Thibault, M.-P., Helloin, M.-C., & Croteau, B. (2003). Exalang – 5/8. Une batterie d'examen du langage oral et écrit chez l'enfant de 5 à 8 ans. *Travaux Neuchâtelois de Linguistique* **38**/**39**, 129–152.

Thibault, M.-P., Lenfant, M., & Helloin, M.-C. (2012). *Bilan informatisé pour l'examen du langage et des compétences transversales chez l'enfant de 8 à 11 ans (ou scolarisé du CE2 au CM2)*. Orthomotus.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). Comprehensive Test of Phonological Processing (CTOPP*). PRO-ED.