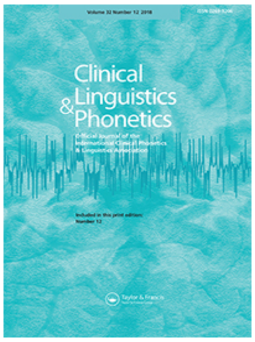

Taylor & Francis
Taylor & Francis Group

Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: http://www.tandfonline.com/loi/iclp20

The relation between phonological and lexical development in French-speaking children

Margaret Kehoe, Tamara Patrucco-Nanchen, Margaret Friend & Pascal Zesiger

To cite this article: Margaret Kehoe, Tamara Patrucco-Nanchen, Margaret Friend & Pascal Zesiger (2018) The relation between phonological and lexical development in French-speaking children, Clinical Linguistics & Phonetics, 32:12, 1103-1125, DOI: 10.1080/02699206.2018.1510984

To link to this article: https://doi.org/10.1080/02699206.2018.1510984

Published online: 05 Sep 2018.

CLINICAL LINGUISTICS & PHONETICS
2018, VOL. 32, NO. 12, 1103–1125
https://doi.org/10.1080/02699206.2018.1510984

The relation between phonological and lexical development in French-speaking children

Margaret Kehoe[a], Tamara Patrucco-Nanchen[a], Margaret Friend[b], and Pascal Zesiger[a]

[a]Department of Psycholinguistics, University of Geneva, Geneva, Switzerland; [b]Department of Psychology, San Diego State University, California, USA

ABSTRACT

This study examines the relation between lexical and phonological variables in 40 French-speaking children, aged 2;5. Specifically, it examines the influence of phonetic complexity, phonological production, phonological memory and neighbourhood density (ND) on vocabulary size. Children were divided into four groups on the basis of their scores on the French version of the Communicative Developmental Inventory (CDI): late1 (< 10%ile), late2 (15–25%ile), middle (40–60%ile) and precocious (> 90%ile). The children's lexicons were coded in terms of phonetic complexity and ND (one-and two-syllable words), and their production capacities were determined from measuring percent consonants correct (PCC) and the number of syllable-initial (C_{SI}) and -final (C_{SF}) consonants in their phonetic inventories. The children also took part in a non-word repetition (NWR) task. Results indicated significant group differences in all four sets of variables. Children with larger vocabularies selected words with greater phonetic complexity and with lower ND values. They had superior PCC, C_{SI} and NWR scores compared to children with smaller vocabularies. Linear regression analyses indicated that 76% of variance in vocabulary size could be accounted for by ND in combination with phonetic complexity and C_{SI}. Our findings are consistent with previous studies which show that ND plays an important role in accounting for variance in vocabulary size. They also indicate that phonetic complexity and phonological production influence lexical acquisition.

ARTICLE HISTORY
Received 3 March 2018
Accepted 6 August 2018

KEYWORDS
French; phonological development; lexical development; children

The relationship between lexical and phonological development has been the subject of much research in recent times (see Stoel-Gammon, 2011, for a review). One line of research has focused on children's tendency to select and avoid words on the basis of their phonological characteristics, a phenomenon referred to as *lexical selection and avoidance* (Ferguson & Farwell, 1975; Schwartz & Leonard, 1982). Another line of research has examined the relation between phonological production and vocabulary size. In particular, studies have examined the phonological abilities of children who have exceptionally small (i.e. late talkers) or large vocabularies (i.e. precocious talkers) (Bortolini & Leonard, 2000; Paul & Jennings, 1992; Petinou & Okalidou, 2006; Rescorla & Ratner, 1996; Smith, McGregor, & Demille, 2006). Allied

CONTACT Margaret Kehoe Margaret.Winkler-Kehoe@unige.ch Dèpartement de psycholinguistique, Facultè de psychologie et des sciences de l'èducation, Universitè de Genève, 28 bd du Pont-d'Arve, 1205 Genève, Suisse
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/iclp.

to these studies are those which find that phonological working memory as determined by results on a non-word repetition (NWR) task[1] distinguish children with low and high vocabulary scores (Chiat & Roy, 2007; Hoff, Core, & Bridges, 2008; Stokes & Klee, 2009a). Finally, another line of research has adopted variables from adult psycholinguistics such as neighbourhood density (ND) and word frequency (WF) to determine which factors account for vocabulary development in children (Stokes, Bleses, Basbøll, & Lambertsen, 2012a; Stokes, Kern, & Dos Santos, 2012b).

The aim of the current study is to bring these themes together when studying the phonological and lexical development of French-speaking children, aged 2;5, who vary according to vocabulary size. Specifically, we examine whether children who have small vocabularies differ from children who have medium or large vocabularies in terms of the phonological and lexical characteristics of their lexicons (i.e. phonetic complexity and ND), the phonological precision of their productions and in terms of their phonological working memories. We also investigate how much variance in vocabulary size is accounted for by the combination of these variables.

Lexical selection and avoidance

Observational studies support the idea that children select and avoid words on the basis of their phonological production capacities. This selection and avoidance may take the form of a preference for particular sounds or syllable structures in words (Ferguson & Farwell, 1975). More controlled evidence comes in the form of experimental studies, which show that children learn to produce non-words containing sounds that they can produce more easily than non-words containing sounds that they cannot produce, thus, confirming the link between phonological experience and lexical acquisition (Schwartz & Leonard, 1982).

Beyond the first word period, authors have focused on lexical selection by examining the phonological characteristics of children's lexicons at different ages on the MacArthur Communicative Developmental Inventory (MCDI – Fenson et al., 1993), a parent-based questionnaire. Stoel-Gammon (1998) compared the phonological features of words from the MCDI that reached age of acquisition by 19 months with those that reached age of acquisition between 20 and 30 months. In terms of syllable structure, the proportions of words containing initial and final clusters increased between the two age ranges. There was a decrease in the proportion of stops and an increase in the proportion of fricatives and liquids in word-initial position. Similar types of findings have been reported by Gayraud and Kern (2007) with the French version of the CDI and by Fletcher et al. (2004) with the Cantonese version of the CDI. All of these studies show that those features which are less frequent in the vocabularies of younger children are those which are acquired later in production.

In the current study, we do not look at lexical selection in younger versus older children but rather in children of the same age who vary according to vocabulary size. Kehoe, Chaplin, Mudry, and Friend (2015) observed phonetic selection tendencies in a group of

[1]It is well acknowledged that a NWR task may tap many phonological skills apart from phonological memory including speech perception, phonological representation and articulatory abilities (see Coady & Evans, 2008). It is most commonly known as a measure of phonological memory which is how we refer to it in the current study.

late talkers. One of the ten late talkers did not select any words with initial clusters, four did not select any words with final clusters and all of them selected fewer words with alveo-palatal fricatives than their peers with medium or exceptionally large vocabularies.

Further information on lexical selection comes from a study by Kern and Dos Santos (2016), Kehoe & Patrucco-Nanchen (2017) which examined whether the phonetic complexity of target words in children's vocabularies could explain variance in vocabulary size in French-speaking children, aged 24–30 months. Phonetic complexity refers to a quantitative measure of the complexity of a target word or production based on a set of phonetic/phonological parameters. They employed Jakielski's (2000) Index of Phonetic Complexity (IPC), which provides a complexity point for phonetic properties such as the presence of consonant variegation and consonant clusters. Other measures of phonetic complexity include Stoel-Gammon's (2010) Word Complexity Measure, which assigns complexity points to a slightly different set of parameters including voiced fricatives and affricates and word-initial unstressed syllables. Kern and Dos Santos (2016) found that phonetic complexity accounted for very little additional variance. This result seems surprising given the findings cited above which suggest that the phonetic composition of children's vocabularies changes as children get older (Gayraud & Kern, 2007; Stoel-Gammon, 1998) and that it varies according to vocabulary size (Kehoe et al., 2015). It may be the case, however, that phonetic complexity plays a role but it is not the most important factor compared with other variables. One of the goals of the current study is to determine how much variance in vocabulary size is accounted for by phonetic complexity.

Phonological working memory and phonological production in late and precocious talkers

Many authors note the parallels between learning a new word and repeating a non-word in a repetition task (Coady & Evans, 2008; Stoel-Gammon, 2011). In both cases, the learner is required to form a robust acoustic representation of the underlying speech units. They also need to have knowledge of the articulatory movements necessary to produce the sequence of sounds in the target words. In the repetition task, the learner stores the representation temporarily and then retrieves it, also reflecting phonological working memory processes. According to Hoff et al. (2008), phonological memory is the link between phonological knowledge (interpreted here as phonological representation) and word learning, and the association between phonological representation and memory provides an explanatory mechanism for many of the observed relations between lexical and phonological development.

Phonological memory has been well recognized as a component of word learning in children three years or older (Gathercole, 2006), but its role in early lexical development has only been explored recently due to the difficulties of designing non-word repetition tasks for young children. Several investigators have now shown that NWR tasks can be successfully employed with two year olds (Hoff et al., 2008; Stokes & Klee, 2009a, 2009b). Hoff et al. (2008) found that NWR accuracy was significantly correlated with vocabulary size in a small group of two-year olds, even after partialling out variance due to repetition of real words which presumably reflects the articulatory demands of the repetition task. Studying a large group of two-year olds, Stokes and Klee (2009a) found that results on a NWR task were the strongest predictor of vocabulary scores among a variety of other demographic and behavioural variables, accounting for 36% of variance in vocabulary scores.

In another field of research, studies have concentrated on individual differences in rate of lexical development to support the link between lexical and phonological development. These studies have shown that children with large vocabularies have superior phonological production abilities relative to children with small vocabularies (Paul & Jennings, 1992; Petinou & Okalidou, 2006; Smith et al., 2006; Stoel-Gammon & Dale, 1988). Rescorla and Ratner (1996), for example, found that late talkers vocalized less often, had smaller consonantal and vocalic inventories and employed a more restricted set of syllable shapes than their typically developing peers. At the other end of the spectrum, Smith et al. (2006) found that lexically precocious two-year olds were superior to their age-matched peers in terms of the number of singleton consonants correct and the percentage of final consonants correct. They evidenced fewer phonological processes such as cluster reduction and final consonant deletion.

The association between vocabulary size and phonological production has been observed in a variety of languages including English, Cypriot Greek, Italian and Cantonese (Bortolini & Leonard, 2000; Fletcher et al., 2004; Paul & Jennings, 1992; Petinou & Okalidou, 2006). To date, however, there has been little research on the phonology of late talkers in French. Furthermore, few studies have examined the amount of variance in vocabulary size accounted for by phonological production versus other phonological variables (e.g. phonological memory, phonetic complexity and ND). This is one of the aims of the current study.

Adult-centred psycholinguistic studies

Stoel-Gammon (2011) contrasted two different approaches to examining the association between lexical and phonological development: child- versus adult-centred approaches. In the former, researchers have analysed children's productions or the target words they select in order to link them to vocabulary development. In adult-centred approaches, researchers have borrowed constructs from language processing in adults to examine the role played by lexical and sub-lexical patterns in the ambient language. We are particularly interested in those studies which have focused on the role of ND in accounting for vocabulary size in children.

Neighbourhood density (ND) refers to the number of phonological neighbours of a word whereby a phonological neighbour is a word that differs from another word by substitution, deletion, or addition of a sound in any word position (Luce & Pisoni, 1998). Words which contain many phonological neighbours are said to belong to dense neighbourhoods whereas those which contain few neighbours belong to sparse neighbourhoods. To take an example from French, the word balle "ball" has a high ND value (n = 42) whereas the word chien "dog" has a relatively low one (n = 12).

A series of studies by Stokes and colleagues shows that the variable of ND accounts for an exceptionally high proportion of variance in the vocabulary size of children acquiring English, French and Danish (Stokes, 2010, 2014; Stokes et al., 2012a, 2012b). In all of these studies, they coded the mean ND and word frequency (WF) of one-syllable words appearing in two year old children's lexicons, based on the CDI. The amount of variance accounted for by ND was 39% in Danish-speaking, 47% in English-speaking and 53% in

French-speaking children. In all cases, WF accounted for a small amount of additional variance (English: 14%; French: 9%; Danish: 3%). In this study, we focus on ND and not on WF in our analysis of phonological and lexical associations.

One salient finding from Stokes et al.'s research is that children with small vocabularies select words with high ND values (see also Coady & Aslin, 2003; Storkel, 2004). Stokes et al. (2012b) posited that words from dense neighbourhoods are less taxing on auditory-verbal short term memories than words from sparse neighbourhoods. By virtue of the fact that they share segments (e.g. balle shares neighbours with bol, bulle, belle) they offer a familiar phonetic stream which facilitates word learning. They hypothesize that all children select words with high NDs at the beginning, which reflects statistical learning of phonological detail in early lexical development (Saffran, 2003). Children with low vocabularies, however, continue to adopt this strategy for an extended period, thus, impeding their later word learning. In the current study we seek to confirm Stokes and colleagues' Stokes et al., (2012b) findings that children with small vocabularies have significantly higher ND values than children with large vocabularies, and that ND accounts for a large percentage of the variance in vocabulary size. We also extend Stokes et al.'s research by measuring ND in two-syllable words (ND2). If ND is a robust strategy used by children to acquire words, ND2 values may also be higher in children with small versus large vocabularies.

Current study

This study focuses on phonological and lexical associations in French-speaking children, aged 2;5. The first aim is to examine whether children, separated into groups according to vocabulary size, differ in the phonological and lexical characteristics of their lexicons (phonetic complexity and ND) and in their phonological memory and production abilities. We predict that children with small vocabularies select words with phonetically simpler forms and with higher ND values than children with large vocabularies. We also predict that they will have inferior phonological memory and production skills compared to children with large vocabularies.

The second aim of the study is to examine how much variance in vocabulary size is accounted for by phonetic complexity, ND, phonological memory and production. Based on Stokes et al.'s (2012a, 2012b) studies, we predict that the highest percentage of variance will be accounted for by ND. Nevertheless, given the correlation that is often observed between phonological memory and production on the one hand and lexicon size on the other, we predict that some additional variance will be accounted for by phonological memory and/or production. Given the lack of research on phonetic complexity, we make no specific predictions, although the findings of Kern and Dos Santos (2016, 2017) lead us to predict that phonetic complexity does not play a strong role in accounting for vocabulary size.

Method

Participants

Participants include 40 monolingual French-speaking children, aged 2;5 (± 15 days). Children were selected from the larger data-base on the basis of their percentile scores on the L'Inventaire Français du Développement Communicatif (IFDC) (Kern & Gayraud, 2010) (the French adaptation of the CDI). Four groups were formed: (1) late1 (n = 8; 3 girls) obtained IFDC scores at or below the 10th percentile (range = 40–221 words); (2) late2 (n = 9; 6 girls) obtained scores between the 15th and 25th percentile (range = 268–353); (3) middle (n = 11; 5 girls) obtained scores between the 40th and 60th percentile (range = 372–474); and (4) precocious (n = 12; 6 girls) obtained scores which exceeded the 90th percentile (range = 572–677). All children had normal hearing, were reported to be in good health and were developing normally.

Procedure

Children attended a single session of 60 minutes in the speech laboratory at the University of Geneva in which they received a battery of tests designed to measure executive function, receptive and expressive vocabulary, and morphosyntax. They were also administered a NWR task. In addition, they engaged in a play session (of 20 minutes duration) with Fisher Price farm toys while interacting with one of their parents. The play items were the same for each child, thus, ensuring a uniform set of vocabulary items per child. The language samples were recorded using a portable digital tape-recorder (Marantz PMD620). The parents completed the IFDC at or prior to the session.

We provide information on the NWR task directly below and information on the analyses of phonetic complexity, ND and phonological production in the section on Data-Coding.

Non-word repetition

The NWR task of Hoff et al. (2008) was modified to make it appropriate for French-speaking participants. It included two training trials with monosyllabic non-words and 12 test trials with one-, two- and three-syllable non-words (four trials per group). The non-words were created using words of the "mots et gestes" form of the IFDC. They were transformed following two basic principles: (1) monosyllabic non-words were created by changing the first phoneme (e.g. dame /**d**am/ "lady" → /bam/); (2) multi-syllabic non-words were created by combining syllables of words from the IFDC which occurred in the same word position (e.g. maman /**ma**mɑ̃/ "mummy" and ballon /ba**lɔ̃**/ "balloon/ (foot) ball" →/malɔ̃/). Appendix A provides a list of the stimuli in the NWR task.

The trials were presented orally by a native French-speaking examiner sitting in front of the children. The non-words were accompanied by toys representing people and animals. During the test, the examiner showed a toy, said the non-word as if it was the toy's name and asked the child to repeat it back (e.g. "This guy is named Bam. Can you say *Bam*?"). If the child didn't repeat the name, the examiner repeated the non-word up to three times.

Only the first repetition produced by the child was scored, regardless of its accuracy. If a child failed to repeat the non-word for six consecutive trials, the test was ended. Only children who attempted to repeat at least three non-words were included in the analyses. The accuracy of NWR was measured by calculating the total number of consonants presented that were repeated correctly by the child. Four of the 40 children did not successfully complete the task. Two of the children belonged to the late1 group and two to the precocious group. The scoring of repetition accuracy was done by French native speakers. The NWR data of ten children (25% of the data) were coded by a second native coder. Phoneme-by-phoneme analysis yielded inter-rater agreement of .97.

Data-coding

Phonetic complexity

A reduced set of the L'Inventaire Français du Développement Communicatif (IFDC-Reduced or IFDC-R) was coded for phonetic complexity. The reduced set included 12 categories of items considered representative of core vocabulary. It omitted context-based items (e.g. people and function words). As well, all words that appeared twice in the inventory (e.g. eau "water", parc "park", pot "pot", poisson "fish", etc.) were limited to one occurrence (Stokes et al., 2012b).

Using the IPC (Jakielski, 2000), a point was assigned to each word of the IFDC-R if it contained: a dorsal consonant (e.g. camion [**k**amjɔ̃] "truck), a fricative or liquid (e.g. avion [a**v**jɔ̃] "plane"; balle [ba**l**] "ball"), a final consonant (e.g. balle [ba**l**] "ball"), three or more syllables (e.g. animal [**animal**] "animal"), two or more consonants with different places of articulation (PoA) (e.g. balle [**b**a**l**] "ball" which has labial and coronal PoAs), a tautosyllabic cluster (e.g. crayon [**kʁ**ejɔ̃] "pencil"), or a heterosyllabic cluster (e.g. tracteur [tʁa**kt**œːʁ]"tractor"). The IPC also assigns points to rhotic vowels but since rhotic vowels do not occur in French, this category was excluded. Once coding was completed, we determined the mean phonetic complexity value for each child based on their lexicon at 2;5. The mean value provides a quantitative measure of the phonetic complexity of children with different-sized vocabularies.

We also wanted to know whether the phonetic complexity of children with small or large vocabularies differs in qualitative ways. To assess this, we calculated the proportion of total phonetic complexity represented by the different components of the IPC. For example, in the target form of the IFDC-R, 17% complexity is represented by dorsals, 32% by fricatives and liquids, 13% by word-final consonants, 5% by multi-syllabic words, 20% by PoA variegation, 10% by tautosyllabic clusters and 2.5% by heterosyllabic clusters. By calculating the proportions of total phonetic complexity for the different components of the IPC per child, we aimed to determine whether vocabulary size differentially influenced certain phonetic domains (e.g. dorsals, fricatives, etc.) more than others.

Neighbourhood density

One and two-syllable words of the IFDC-R were coded for ND using the values generated by the Lexique3 database, a corpus of adult language (New, Brysbaert, Veronis, & Pallier, 2007). In the case of verbs, we followed the procedure of Stokes et al. (2012b) by coding the most frequent phonological form. In some cases, the infinitive form was coded (e.g. laver "to wash"), whereas in other cases, a morphologically derived form was coded (e.g. donne "give" for the verb donner "to give"). Similarly, the most frequent phonological forms were chosen when two noun or two adjective choices were provided (e.g. **figure**/visage "face"; beau/**belle** "beautiful"). Once coding was completed, a mean ND value was obtained separately for one- and two-syllable words (ND1, ND2) for each child's lexicon at 2;5.[2]

[2]Unlike Stokes et al. (2012b), we do not include two syllable words such as "p(e)tit" and "ch(e)val" in the analyses of ND for one-syllable words. Although these words may be realized as monosyllabic, research by Andreassen (2013) shows that the variant containing schwa is more frequent in children's productions as well as in the input to children. Thus, these words have been coded as disyllabic.

Phonological production

We analyzed children's spontaneous language samples using Phon, a software program specifically designed for the analysis of phonological data (Rose et al., 2006). Each child's wave file was segmented into utterances, glossed and phonetically transcribed. Two French-speaking graduate students, who had experience in phonetic transcription, performed the analyses. Three measures of phonological production were obtained: Percent consonants correct (PCC) and the number of consonants in syllable-initial (C_{SI}) and -final position (C_{SF}) in the children's phonetic inventories. A consonant was designated as being part of the phonetic inventory if it was present at least two times and in two different words. Calculations of PCC were computed automatically for each child based on the entire number of utterances in the 20 minute recordings using the query function PCC-PVC in Phon. Three participants were re-transcribed by a second transcriber using the Blind Transcription function of Phon. Point-to-point agreement in terms of consonant transcription was high (ranging from 88% to 93%).

Statistical analyses

The statistical analyses were performed using R statistical software (R Development Core Team, 2016). Group differences were determined using one-way analyses of variance (ANOVA) and Tukey post-hoc tests. Due to the number of multiple comparisons being made and the possibility of inflating type 1 errors, we adopted the critical *p* value of 0.01 which was close to the Bonferroni correction (.05/6) of 0.008. The amount of variance in vocabulary size explained by the predictor variables was determined by linear regression (lm function in R).

Results

Phonetic complexity

Figure 1 shows a box-plot representation of mean phonetic complexity values for the French-speaking children separated according to vocabulary size. Mean phonetic complexity ranged from 2.83 through to 4.07. A one-way ANOVA revealed a significant group effect ($F(3,36) = 10.87$, $p < 0.001$). Children with larger vocabularies had higher phonetic complexity values than children with small vocabularies. Tukey HSD multiple comparisons indicated that the precocious group differed significantly from late1 ($p < 0.001$), late2 ($p < 0.01$) and middle groups ($p < 0.01$) but there were no significant differences between the two late and middle groups.[3]

To examine whether there were qualitative differences in phonetic complexity between the lexicons of late1, late2, middle and precocious children, we calculated the proportion of total complexity occupied by the different sub-components of the IPC for each group of children. Mean values per group are shown in Figure 2 along with the values based on the IFDC-R ($n = 518$ items). Figure 2 indicates that phonetic complexity was realized in similar proportions by all children with the bulk of complexity going to the categories of fricatives/liquids, 2+ PoA, dorsals

[3]One reason for the lack of significant differences in the late groups was the reduced numbers of words in their lexicons. Child 66 (late1), for example, obtained the highest phonetic complexity value (i.e. 4.07) of all children but his score was based on only 15 items, several of which were high phonetic complexity words (e.g. camion de pompier, glace, chocolat, popcorn).

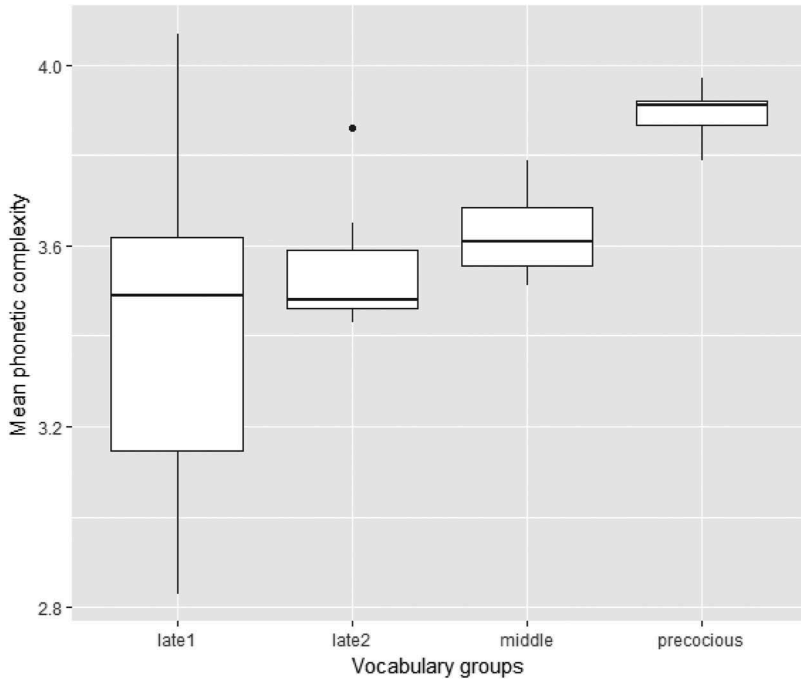

Figure 1. Box plot representations of mean phonetic complexity in the four groups of French-speaking children separated according to vocabulary size. In a boxplot display, the center line represents the median (50th percentile), the bottom and top of the box, the 25th and 75th percentile, and the whiskers, the minimum and maximum values. Outliers are shown as individual points.

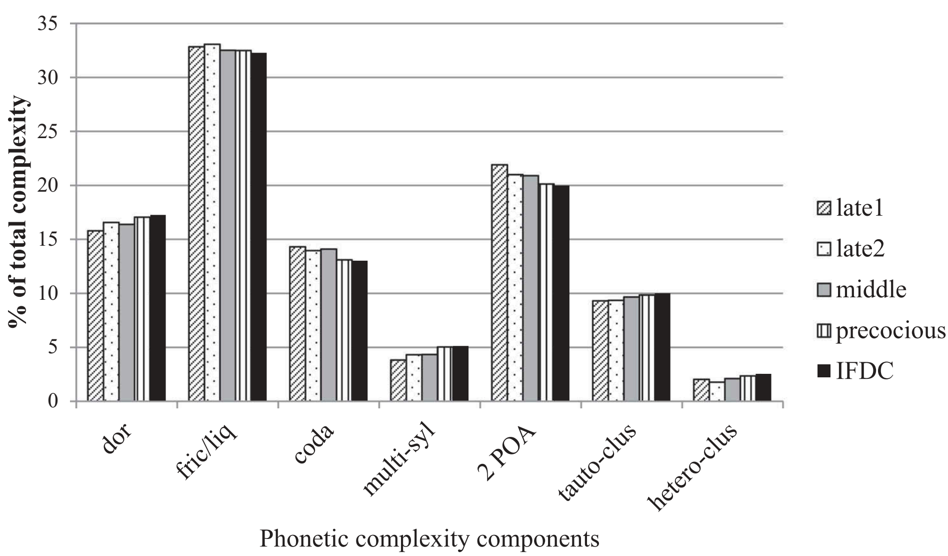

Figure 2. Percentage of total phonetic complexity represented by the different phonetic components of the IPC for the four vocabulary groups and for the restricted target set of the IFDC.

and final codas. Nevertheless, there were subtle differences amongst the four vocabulary groups in terms of their realization of phonetic complexity. The late talkers selected fewer words with dorsals, three-syllables or more, and with tauto- and hetero-syllabic clusters. They compensated for their reduced complexity in these areas by producing greater complexity in the categories of fricatives/liquids, codas and multiple PoAs. This latter finding does not mean that they produced more words with fricatives/liquids, codas and multiple PoAs than children with larger vocabularies; rather, it means that these were the areas of complexity which posed less difficulty for them.

One-way ANOVAs indicated that there were significant group differences for three of the seven IPC sub-components: final codas ($F(3,36) = 9.58$, $p < 0.001$), multi-syllabic words (($F(3,36) = 9.10$, $p < 0.001$) and 2+ PoA ($F(3,36) = 14.30$, $p < 0.001$). Tukey HSD multiple comparisons indicated that precocious children produced more multi-syllabic words than the late1 group ($p < 0.001$). They produced fewer words with final codas (percentage-wise as a proportion of total complexity) than the late1 ($p < 0.001$), late2 ($p < 0.01$) and middle ($p < 0.001$) groups and fewer words with 2+ PoA than the late1 group ($p < 0.001$). The middle group also differed from the late1 group in 2+ PoA ($p < 0.01$). In sum, the findings provide support for different lexical selection patterns amongst the four vocabulary groups.

Phonological memory

Figure 3 displays the NWR scores of the four groups of children separated according to vocabulary size. NWR scores ranged from 1 through to 32 with a mean of 21 (out of a total of

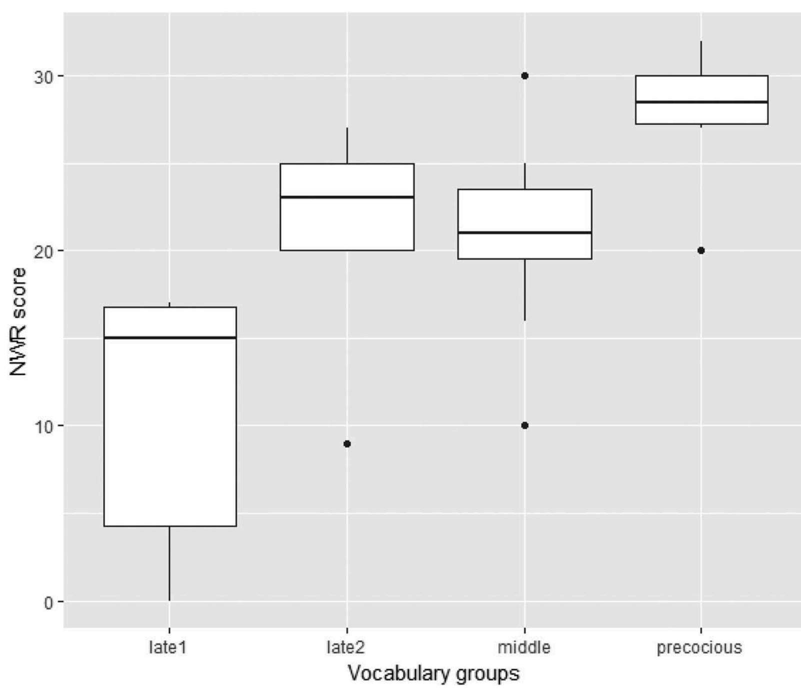

Figure 3. Box plot representations of NWR scores in the four groups of French-speaking children separated according to vocabulary size.

32). Our statistical analyses here are based on 36 rather than 40 children since four children did not complete the task. A one-way analysis of variance (ANOVA) revealed a significant group effect ($F(3,32) = 11.28$, $p < 0.001$). Children with larger vocabularies had higher NWR scores than children with small vocabularies. Tukey HSD multiple comparisons indicated that the precocious and middle groups differed significantly from the late1 group ($p < 0.001$; $p < 0.01$). There were no significant differences between the late2 and middle groups.

Phonological production

The average PCC value for the French-speaking children was 79.94% and ranged from 42.81% to 95.45%. The PCC results are shown in Figure 4 for the four different vocabulary groups. A one-way ANOVA revealed a significant group effect ($F(3,36) = 7.34$, $p < 0.001$). Children with larger vocabularies had superior PCCs than children with smaller vocabularies. Tukey HSD multiple comparisons indicated that the precocious group differed significantly from the late1 ($p < 0.001$) group. There were no significant differences between the other groups.

The average number of consonants in syllable-initial and -final position in the children's phonetic inventories was 16 (range: 6–19) and 5 (range: 1–9) respectively. The phonetic inventory results for syllable-initial and -final position across the four groups of children are shown in Figures 5 and 6 respectively. A one-way ANOVA indicated a significant group effect for C_{SI} ($F(3,36) = 9.04$, $p < 0.001$). Children with

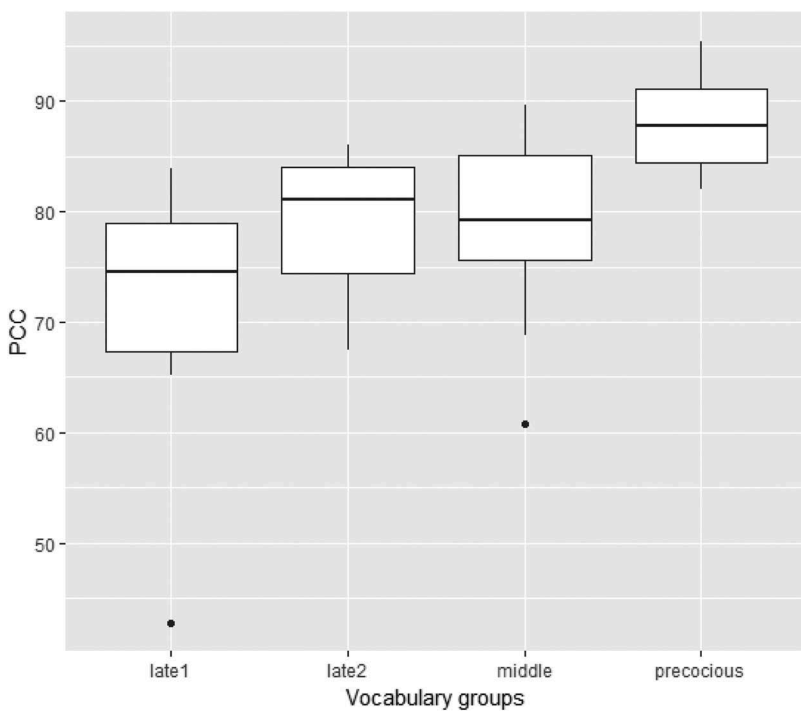

Figure 4. Box plot representations of PCC in the four groups of French-speaking children separated according to vocabulary size.

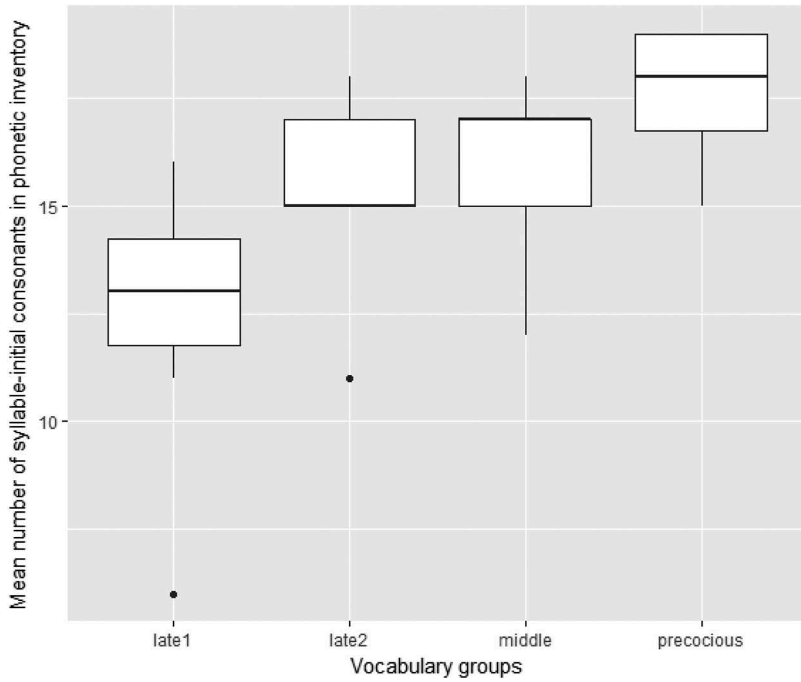

Figure 5. Box plot representations of mean number of syllable-initial consonants in the phonetic inventories of the four groups of French-speaking children separated according to vocabulary size. In the late2 group, the median and lower quartile values were the same (i.e. 15); in the middle group, the median and the upper quartile values (i.e. 18) were the same.

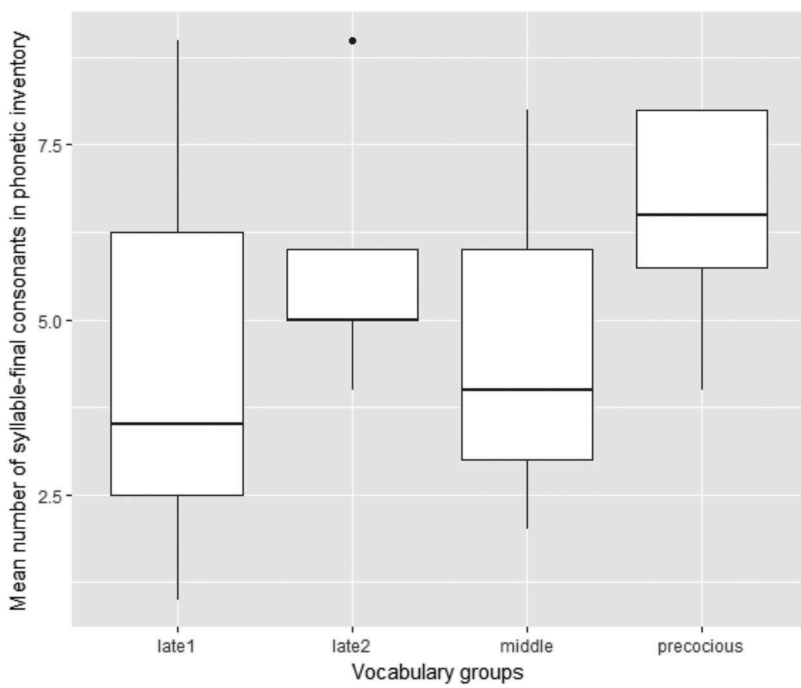

Figure 6. Box plot representations of mean number of syllable-final consonants in the phonetic inventories of the four groups of French-speaking children separated according to vocabulary size. In the late2 group, the median and the lower quartile values were the same (i.e. 5.0).

larger vocabularies had greater numbers of C_{SI} than children with smaller vocabularies. Tukey HSD multiple comparisons indicated that the precocious and middle groups differed significantly from the late1 group ($p < 0.001$; $p < 0.01$). There were no significant differences between the other groups. As for C_{SF}, a one-way ANOVA indicated a significant group effect ($F(3,36) = 3.01$, $p < 0.05$) but there were no significant differences between groups when follow-up multiple comparisons were conducted.

Neighbourhood density

ND values were calculated separately for one- (ND1) and two-syllable words (ND2). The average ND1 was 24.00 and ranged from 22.40 through to 27.57. The average ND2 was 10.26 and ranged from 7.95 through to 12.15. The ND1 and ND2 results are shown in Figures 7 and 8 for the four groups of French-speaking children. A one-way ANOVA revealed a significant group effect for ND1 ($F(3,36) = 11.02$, $p < 0.001$). Children with smaller vocabularies had higher ND1 values than children with larger vocabularies. Tukey HSD multiple comparisons indicated that the precocious group differed significantly from late1 ($p < 0.001$) and late2 ($p < 0.01$) but not from the middle group. The middle group differed significantly from the late1 ($p < 0.01$) but not from the late2 group. The two late groups did not differ significantly from each other. In the case of ND2, a one-way ANOVA indicated no significant difference between groups ($F(3,36) = 1.37$, $p > 0.05$).

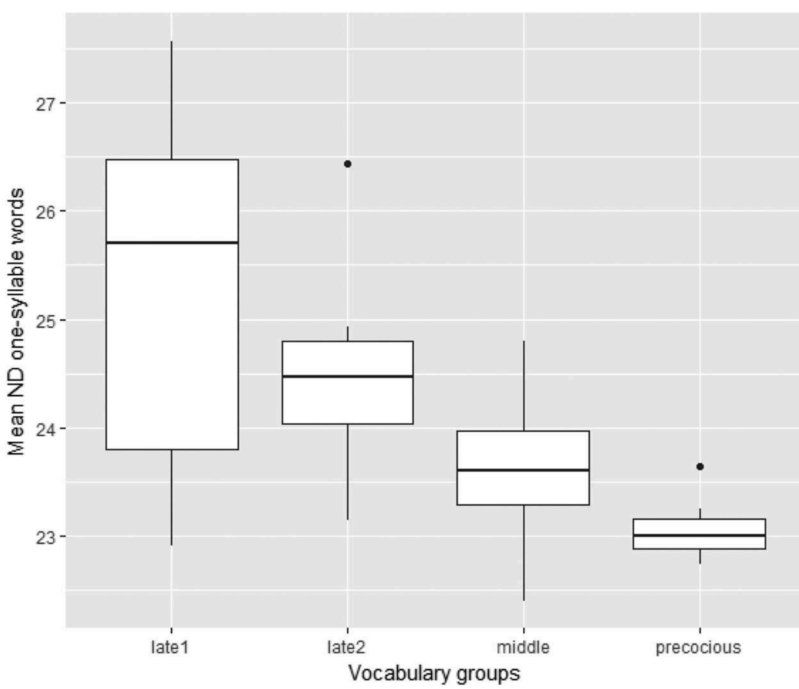

Figure 7. Box plot representations of ND for one-syllable words in the four groups of French-speaking children separated according to vocabulary size.

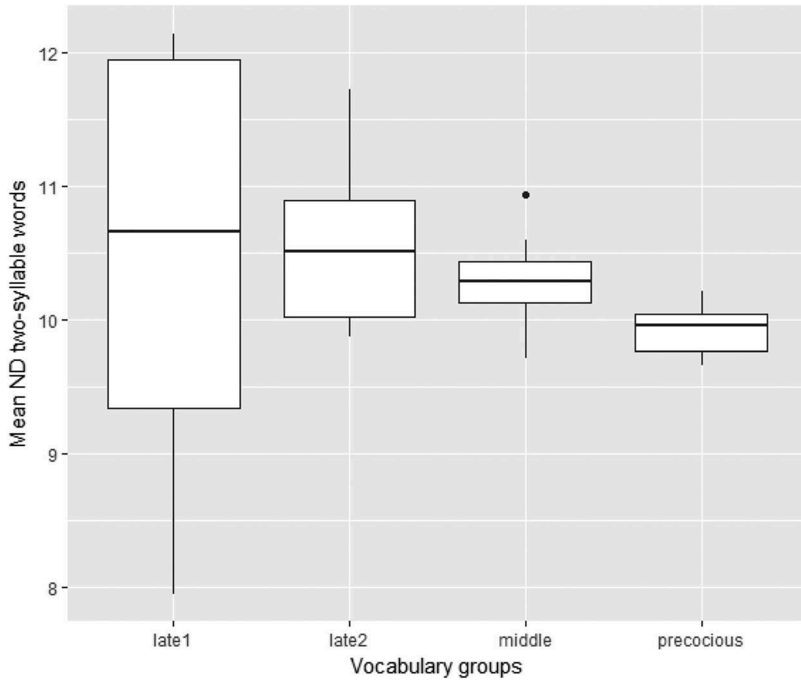

Figure 8. Box plot representations of ND for two-syllable words in the four groups of French-speaking children separated according to vocabulary size.

Correlations and linear regression

Table 1 displays the correlations amongst the different lexical and phonological variables. Moderate to moderately-high significant correlations were observed between all variables with the exception of ND2 which was moderately correlated with phonetic complexity and ND1 only, and phonetic complexity which was not correlated with phonological memory or with the production variables based on the phonetic inventory. Vocabulary size had moderately high correlations with phonetic complexity, phonological memory, phonological production (in particular, C_{SI}) and ND1.

Table 1. Correlations amongst lexical and phonological variables.

Variables	1	2	3	4	5	6	7	8
1. Vocab size (IFDC)	_	.67***	.76***	.62***	.69***	.40*	–.76***	–.33*
2. Phonetic complexity		_	.30	.58***	.24	.25	–.48**	–.61***
3. Phonological memory (NWR)			_	.53**	.75***	.48**	–.61***	.09
4. Phonological production (PCC)				_	.50**	.48**	–.48**	–.18
5. Phonological production (C_{SI})					_	.52**	–.68***	–.12
6. Phonological production (C_{SF})						_	–.50**	–.18
7. ND1							_	.46**
8. ND2								_

*$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

We conducted a linear regression model entering the seven variables, phonetic complexity, NWR, PCC, C_{SI}, C_{SF}, ND1 and ND2 as predictor variables and using total vocabulary size at 2;5 as output variable.[4] Due to the fact that there were missing data for NWR, we used the impute function in R to provide five different estimates of missing values and, accordingly, to compute five different models based on the estimates. We then used the pooled model function to determine which variables were significant in our model. It showed that three variables were significant: Phonetic complexity, ND1 and C_{SI}.[5] We then entered the three variables individually to determine which accounted for the most variance. The results were as follows: ND1 accounted for the highest unique variance (57%), followed by phonetic complexity (an additional 11% unique variance) and C_{SI} (an additional 8% unique variance). The final model accounted for 76% of the variance of total vocabulary size. It is shown in Table 2.

Discussion

The purpose of the study was to examine the influence of phonological and lexical variables on the vocabulary development of French-speaking children, aged 2;5. We aimed to determine whether children who have small vocabularies differ from children who have medium and large vocabularies in terms of the phonetic complexity and neighbourhood density (ND) of their lexicons and in terms of their phonological memory and production. We also aimed to determine how much variance in vocabulary size could be accounted for by these variables. Our results showed significant group differences in all four sets of variables. They also showed that a high proportion of variance in total vocabulary size could be accounted for by ND in combination with phonetic complexity and phonological production. In the following paragraphs, we discuss the findings in more detail.

Table 2. Results of hierarchical regression analyses for predicting vocabulary size.

Model	*β*	SE	t	*p* value
Model 1: $IFDC_{tot}$ ~ ND1+ Phonetic complexity + C_{SI}				
1. Intercept	3095.86	374.54	8.266	< 0.001
ND1	−112.29	15.59	−7.201	< 0.001
2. Intercept	1377.25	545.77	2.524	< 0.05
ND1	−84.13	15.16	−5.549	< 0.001
Phonetic complexity	285.98	73.59	3.886	< 0.001
3. Intercept	−131.890	636.216	−.207	> 0.05
ND1	−42.664	17.590	−2.426	< 0.05
Phonetic complexity	316.468	64.677	4.893	< 0.001
C_{SI}	25.765	7.217	3.570	< 0.01

[4]We use the total IFDC score as our dependent variable of vocabulary size. Preliminary analyses revealed that similar correlations were obtained regardless of whether we took the total IFDC or the IFDC-R as dependent variable.

[5]The Variance Inflation Factors (VIF) were 2.30 for ND1, 1.32 for phonetic complexity and 1.9 for C_{SI} suggesting acceptable levels for multi-collinearity.

Group differences

The children in this study were divided into four groups based on their vocabularies sizes at 2;5. Our results indicated that children in the precocious group selected words with greater phonetic complexity and with lower ND than children in the late and middle groups. They had superior phonological working memories and phonological production abilities. Multiple comparisons did not reveal significant differences between all four groups however. The precocious group, which was characterized by low intra-group variability, tended to differ from the other groups, whereas the late1 group, which was characterized by high intra-group variability, tended to have similar scores to the late2 and middle groups. The lack of significant group effects for the low vocabulary groups may result from reduced power relating to sampling effects: the number of items was reduced in the late talkers in comparison to the other groups.

These findings are consistent with those of Stokes and colleagues which show that children with low vocabulary sizes select words from high density neighbourhoods (Stokes et al., 2012a, 2012b). They studied a group of children ranging in age from 2;0 to 2;6 whereas we focused on children, aged 2;5 only. Thus our results indicate that even at the outer limits of the age range 2;0 to 2;6, the effects of ND appear to be strong. We extended Stokes et al.'s (2012a, 2012b) research by also examining ND for two-syllable words. We hypothesized that if ND is a robust strategy for word learning, its effects may also be evident in the analyses of longer words. In our study, one-syllable words represented only 38% of children's vocabularies, whereas two-syllable words represented 44%. Although there was a clear trend for the ND of two-syllable words to decline as vocabulary size became smaller (see Figure 8), our statistical tests did not show a significant group effect, nor did this measure correlate with other lexical and phonological variables to the same degree as the ND of one-syllable words (see Table 1). Thus, we conclude that the effects of ND on word learning pertain particularly to one-syllable words.

Our results are consistent with numerous studies showing that late talkers have inferior and precocious talkers have superior production abilities compared to their typically developing peers (Paul & Jennings, 1992; Rescorla & Ratner, 1996; Smith et al., 2006). These studies have used a variety of measures to tap phonological production including phonetic inventories, consonant and vowel precision (PCC and PVC), and the number of phonological processes present in spontaneous speech. In this study, we utilized three production measures, PCC, C_{SI} and C_{SF}, which we extracted from phonetic transcriptions of the children's spontaneous speech. PCC and C_{SI} proved sensitive to group differences in vocabulary size whereas C_{SF} did not. Previous analyses on a similar database of children also found phonological measures based on syllable-initial position to be more sensitive to lexical effects than those based on syllable-final position (Kehoe et al., 2015). These results are in opposition to those conducted on English-speaking children, which find stronger correlations between vocabulary size and phonological measures based on syllable-final position (Rescorla & Ratner, 1996; Smith et al., 2006). We assume that differences in the phonological structure of English and French underlie these different effects, French having a higher proportion of open syllables than English (Delattre & Olsen, 1969). As vocabulary size grows, English-speaking children need to represent and produce final consonants accurately in order to maintain functional differences between words. In contrast, French-speaking children need to do this to a lesser extent. Instead, they need to concentrate on syllable-initial contrasts.

Similarly, our findings are in accordance with numerous studies which show a relationship between vocabulary knowledge and phonological working memory in children as young as two years (Hoff et al., 2008; Stokes & Klee, 2009b). Children with small vocabularies may be less efficient at supporting short-term memory representations or creating long-term memory representations of phonemes due to their reduced lexical knowledge (Archibald & Gathercole, 2006). Conversely, their poor phonological memory skills may make them less able to learn new words (Hoff et al., 2008).

There has been less research on the phonetic complexity of words selected by late and precocious talkers. Nevertheless, our findings support studies which have examined the phonological characteristics of children's lexicons at different ages (Fletcher et al., 2004; Gayraud & Kern, 2007; Stoel-Gammon, 1998). These studies show that older children's lexicons contain phonetically more complex words than younger children's. Our results indicate that the lexicons of lexically advanced children contain phonetically more complex words than the lexicons of lexically impoverished children.

The phonetic complexity values of the late1 group were also characterized by extreme variability (see Figure 1). As mentioned, some of this variability may arise from methodological factors related to the reduced numbers of items used to compute phonetic complexity for some of the late talkers. However, some of this variability may signal important individual differences with implications for later development. The late talkers who choose words with high phonetic complexity may resolve their expressive language difficulties whereas those who choose words with low phonetic complexity may present with expressive language difficulties later on. Presumably, choosing phonetically more complex words leads to higher quality phonological representations which in turn leads to enhanced word learning. Note that Stokes et al. (2012b) made a similar type of prediction concerning individual differences in ND values amongst late talkers. They hypothesized that those late talkers who have ND values similar to their typically developing peers may have failed to use a statistical learning mechanism, restricting learning to highly functional words. They may have less difficulty proceeding to later stages of word learning which require mapping of words from sparser neighbourhoods to semantic representations. A longitudinal study design would be important to confirm whether phonetic complexity and ND values at 2;5 are predictive of later vocabulary development.

Our analyses of phonetic complexity also focused on the individual components of the IPC. We observed relative differences between vocabulary groups across phonetic parameters leading us to infer that words containing dorsal consonants, three or more syllables, tauto- (e.g. **tr**acteur) and hetero-syllabic clusters (e.g. tra**ct**eur) are likely to be selected less often by children with low vocabularies than words containing other phonetic parameters (e.g. fricatives/liquids, word-final codas, 2+ PoA). In essence, these parameters were the phonetically most complex of the IPC. Statistical analyses confirmed that there were significant differences amongst the vocab groups, suggesting that phonetic properties play a role in children's selection of words.

Regression analyses

The linear regression analyses revealed that the ND of one-syllable words, phonetic complexity and the production variable, C_{SI}, accounted for 76% variance in vocabulary size. Nevertheless, the bulk of the effect was carried by ND1. This result mirrors work by Stokes and colleagues which indicates that ND alone accounts for a surprisingly high proportion of variance in children's vocabulary development (Stokes et al., 2012b). That is, the number of phonological neighbours of one-syllable words in children's lexicons is highly predictive of children's overall vocabulary size.

Our study design differs from Stokes and colleagues in that we did not include word frequency (WF) as variable (Stokes, 2010; Stokes et al., 2012a, 2012b). They found that it played a reduced role in their regression models in comparison to ND, and, thus, we chose to focus on other phonological variables such as phonetic complexity and production. More recently, Kern and Dos Santos (2017) have reanalysed Stokes et al.'s (2012b) French data and found WF to play a stronger role. They separated out one-syllable words according to grammatical category (nouns vs. predicates) and used median rather than mean scores for WF. The latter was an attempt to neutralize the effects of extreme values that arose from the inclusion of high frequency verbs. Kern and Dos Santos (2017) found a similar pattern of results to Stokes et al. (2012b) when nouns and predicates were grouped together but when they separated out grammatical category, WF accounted for more variance in vocabulary size for nouns than ND did, and neither of them accounted for any variance in vocabulary size for predicates. Thus, it is possible that a reanalysis of the data including the variable of WF may yield different results to the current ones, particularly if grammatical class is taken into consideration.

Kern and Dos Santos (2017) also looked at phonetic complexity but found it not to account for any additional variance beyond ND and WF. Our study differs from theirs in that we calculated phonetic complexity across the entire IFDC-R rather than only for one-syllable words. In preliminary analyses, we found the phonetic complexity of one-syllable words to be minimally correlated with vocabulary size (r = .37) but the phonetic complexity of the entire vocabulary set to be moderately to highly correlated (r = .67), possibly because certain components of the IPC (i.e. presence of multi-syllabic words and hetero-syllabic clusters) cannot be indexed in one-syllable words. The current results show that phonetic complexity when based on the IFDC-R explains additional variance to ND1, suggesting that children with low vocabularies are seeking phonetically simpler words regardless of the neighbourhoods in which these words reside.

Maekawa and Storkel (2006) found that ND and WF accounted for word learning, specifically, the age of first production of words, but it did so for only one of the three children studied. The variable that affected word learning for all three children was word length, a variable that takes into consideration the number of phonemes in the target word. Given that this variable also measures complexity of the target word, it resembles our measure of phonetic complexity. Indeed, the IPC assigns points to multi-syllabic words and words with final consonants and clusters, and, as such, measures word length indirectly. Future studies should determine whether phonetic complexity accounts for similar amounts of variance as word length or whether it is even more sensitive, since it considers phonological features which pose difficulty in production such as manner (e.g. fricatives/liquids) and place (e.g. dorsals, 2+ PoA) of articulation.

One of the most important findings of the present study is that phonological production matters to vocabulary learning. The association between phonological production and lexicon size is well established, but few studies have examined the proportion of variance accounted for by production versus other phonological or lexical variables. In an earlier analysis, PCC accounted for a small amount of unique variance (2%) over and above ND and phonetic complexity (Kehoe & Patrucco-Nanchen, 2017). In the current analysis, C_{SI} accounted for a greater amount of unique variance (8%) and thus proved to be a more sensitive measure of lexical-based production effects than PCC. Our results suggest that children who have fewer sounds in their phonetic inventories, particularly in their syllable-initial phonetic inventories, are hindered in their vocabulary learning.

Recently, Zamuner and Thiessen (2018), show that children's production experience with the sounds of a target word accounts for variance in their likelihood of imitating that target word. The model that best accounted for imitation contained variables related to the properties of the target word (i.e. ND) and the child's production experience, results not dissimilar to ours. These findings in conjunction with previous work in lexical selection (Ferguson & Farwell, 1975) and on the "articulatory filter" (Vihman, 1996, 2017) strongly suggest that words containing sounds that children can produce will be more salient to them, make fewer processing demands, be more robustly represented in memory, and, consequently, will be more likely to be learned. Of course, the causal direction between phonological production and the lexicon cannot be resolved from the regression findings and, thus, an alternative interpretation of the results is that children who have small vocabularies are prevented from developing their phonological production capacities. Current models of lexical-phonological interactions which emphasize the dynamic relationship between the two domains would lead us to posit that word learning and phonological development occur in tandem (Edwards, Munson, & Beckman, 2011; Stoel-Gammon, 2011).

Phonological working memory did not prove significant in our regression analysis, although its effect may have been weakened by the need to impute missing data. Hoff et al. (2008) found a relationship between phonological working memory and vocabulary size in two-year olds, after partialling out the variance shared with real-word repetition. They argued that this finding was evidence that phonological working memory was tapping into "something more" than articulation accuracy, namely, the phonological memory component. In the current study, phonological production, as measured by $C_{SI,}$ emerged as a stronger predictor of vocabulary size than phonological memory. At later age ranges, when consonant inventory sizes are homogeneous and at ceiling levels, phonological memory may prove to be the more sensitive measure of vocabulary development.

Limitations and future directions

This study focused only on lexical and phonological variables and we do not exclude that other types of variables (i.e., demographic, behavioural, social-pragmatic, semantic) may explain variance in vocabulary size more completely than this set of variables. We also do not exclude that other phonological variables such as phonotactic probability may yield similar results to the current ones. Given the high correlations between phonological and lexical variables (Storkel, 2009), it may well be that a single phonological component as yet to be determined is at the root of the significant regression results.

In sum, our findings suggest that several phonological and lexical factors influence children's vocabulary size. They are ND, the phonetic complexity of the target word, and the phonological production skills of the children. Together these factors accounted for an impressive proportion of the variance in children's vocabulary size. Our findings are consistent with the notion that children, aged 2;5, are reliant on ND to support word learning. Equally, we show that phonological production is important to vocabulary acquisition since children with fewer sounds in their phonetic inventories also have smaller vocabularies. In future research, we would like to enter word length and WF into our analyses in order to ensure that the variance in vocabulary size currently accounted for is not better explained by other inter-related variables. In addition, we would like to examine the findings longitudinally to determine which variables among the ones we have measured are the most predictive of children's later vocabulary development.

Acknowledgements

I would like to thank Christophe dos Santos and Sophie Kern for their advice and helpful discussion. I would also like to thank the following students for their assistance in phonetic transcription and data-coding: Lucile da Silva, Elisa Chaplin and Pauline Mudry.

Funding

This study was supported by a grant [#R01HD068458] from the National Institute of Child Health and Human Development awarded to Margaret Friend, Diane Poulin-Dubois and Pascal Zesiger.

Declaration of interest

The authors report no conflicts of interest.

References

Andreassen, H. (2013). *Schwa. distribution and acquisition in light of Swiss French data (Doctoral dissertation)*. Tromsø, Norway: University of Tromsø.

Archibald, L., & Gathercole, S. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*, 970–983. doi:10.1044/1092-4388(2006/070)

Bortolini, U., & Leonard, L. B. (2000). Phonology and children with specific language impairment: Status of structural constraints in two languages. *Journal of Communication Disorders*, *33*, 131–150.

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*, *50*, 429–443. doi:10.1044/1092-4388(2007/030)

Coady, J., & Aslin, R. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*, 441–469.

Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments. *International Journal of Language and Communication Disorders*, *43*, 1–40. doi:10.1080/13682820601116485

Delattre, P., & Olsen, C. (1969). Syllable features and phonic impression in English, German, French, and Spanish. *Lingua. International Review of General Linguistics. Revue Internationale De Linguistique Generale*, *22*, 160–175. doi:10.1016/0024-3841(69)90051-5

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon–Phonology relationships and dynamics of early language development–A commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language*, *38*, 35–40. doi:10.1017/S0305000910000450

Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., … Reilly, J. (1993). *MacArthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Ferguson, C., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439. doi:10.2307/412864

Fletcher, P., Chan, C., Wong, P., Stokes, S., Tardif, T., & Leung, S. (2004). The interface between phonetic and lexical abilities in early Cantonese language development. *Clinical Linguistics & Phonetics*, *18*, 535–545. doi:10.1080/02699200410001703655

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Gayraud, F., & Kern, S. (2007). Caractéristiques phonologiques des noms en fonction de l'âge d'acquisition. *Enfance; Psychologie, Pedagogie, Neuropsychiatrie, Sociologie*, *59*, 324–338. doi:10.3917/enf.594.0324

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development. *Journal of Child Language*, *35*, 903–916. doi:10.1017/S0305000908008751

Jakielski, K. J. (2000, June). *Quantifying phonetic complexity in words: An experimental index*. Paper presented at the Annual Child Phonology Conference, Cedar Falls, IA.

Kehoe, M., & Patrucco-Nanchen, T. (2017). Phonology of late and precocious talkers in French. In E. Baboutsouli. (Ed.), *Proceedings of the international symposium on monolingual and bilingual speech 2017* (pp. 155–161) ISBN 978-618-82351-1-3 URL http://ismbs.eu/publications-2017

Kehoe, M., Chaplin, E., Mudry, P., & Friend, M. (2015). La relation entre le développement du lexique et de la phonologie chez les enfants francophones. *Rééducation Orthophonique*, *263*, 61–85.

Kern, S., & Dos Santos, C. (2017). Invariance in variation. Frequency and neighbourhood density as predictors of vocabulary size. In M. Hickmann, E. Veneziano, & H. Jisa (Eds.), *Sources of variation in first language acquisition. Languages, contexts, and learners. Trends in Language Acquisition Research (TILAR)* (pp. 183–200). Amsterdam, The Netherlands: John Benjamin.

Kern, S., & Dos Santos, C. (2016, September). *Influence of frequency, neighbourhood density and phonetic complexity on first words*. Paper given at the VIII Congresso Internacional de Adquisición del Lenguaje. Palma, Spain.

Kern, S., & Gayraud, F. (2010). *L'inventaire français du développement communicatif*. Grenoble, France: Editions La Cigale.

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1–36. doi:10.1097/00003446-199802000-00001

Maekawa, J., & Storkel, H. (2006). Individual differences in the influence of phonological characteristics on expressive vocabulary development by young children. *Journal of Child Language*, *33*, 439–459.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*, 661–677. doi:10.1017/S014271640707035X

Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech, Language, and Hearing Research*, *35*, 99–107. doi:10.1044/jshr.3501.99

Petinou, K., & Okalidou, A. (2006). Speech patterns in Cypriot-Greek late talkers. *Applied Psycholinguistics*, *27*, 335–353. doi:10.1017/S0142716406060309

R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, *39*, 153–165. doi:10.1044/jshr.3901.153

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 D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Boston University Conference on Language Development* (pp. 489–500). Somerville, MA: Cascadilla Press.

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114. doi:10.1111/1467-8721.01243

Schwartz, R., & Leonard, L. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*, 319–336.

Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics*, *27*, 355–375. doi:10.1017/S0142716406060310

Stoel-Gammon, C. (1998). Sounds and words in early language acquisition. The relationship between lexical and phonological development. In R. Paul (Ed.), *Exploring the speech-language connection* (pp. 25–52). Baltimore, MD: Paul Brookes Publishing.

Stoel-Gammon, C. (2010). The word complexity measure: Description and application to developmental phonology and disorders. *Clinical Linguistics & Phonetics*, *24*, 271–282. doi:10.3109/02699200903581059

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*, 1–34. doi:10.1017/S0305000910000425

Stoel-Gammon, C., & Dale, P. (1988). *Aspects of phonological development of linguistically precocious children*. Paper presented at the Child Phonology Conference, Champaign-Urbana. 10.3168/jds.S0022-0302(88)79586-7

Stokes, S. (2010). Neighborhood density and word frequency in toddlers. *Journal of Speech, Language, and Hearing Research*, *53*, 670–683. doi:10.1044/1092-4388(2009/08-0254)

Stokes, S. (2014). The impact of phonological neighbourhood density on typical and atypical emerging lexicons. *Journal of Child Language*, *41*, 634–657. doi:10.1017/S030500091300010X

Stokes, S., Bleses, D., Basbøll, H., & Lambertsen, C. (2012a). Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech, Language, and Hearing Research*, *55*, 1265–1273. doi:10.1044/1092-4388(2012/10-0291)

Stokes, S., & Klee, T. (2009a). Factors that influence vocabulary development in two-year-old children. *Child Psychology and Psychiatry*, *50*, 498–505. doi:10.1111/j.1469-7610.2008.01991.x

Stokes, S., & Klee, T. (2009b). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language, and Hearing Research*, *52*, 872–882. doi:10.1044/1092-4388(2009/08-0030)

Stokes, S. F., Kern, S., & Dos Santos, C. D. (2012b). Extended Statistical Learning as an account for slow vocabulary growth. *Journal of Child Language*, *39*, 105–129. doi:10.1017/S0305000911000031

Storkel, H. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201–221. doi:10.1017/S0142716404001109

Storkel, H. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, *36*, 291–321. doi:10.1017/S030500090800891X

Vihman, M. (1996). *Phonological development: The origins of language in the child*. Oxford, UK: Basil Blackwell.

Vihman, M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*, 1–27. doi:10.1111/bjop.12207

Zamuner, T., & Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics*. https://doi.org/10.1017/cnj.2018.10.

Appendix A

Stimuli in French NWR task

	One-syllable	Two-syllables	Three-syllables
1	/dyl/	/maˈlɔ̃/	/nupiˈʁœj/
2	/bam/	/gaˈbu/	/pileˈko/
3	/ʁup/	/suˈtœʁ/	/tefiˈlɔ̃/
4	/sɔb/	/fɔˈnɛt/	/sikɔˈmal/