JSLHR

Research Article

The Relationship Between Lexical and Phonological Development in French-Speaking Children: A Longitudinal Study

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

Purpose: This study examines the influence of lexical and phonological factors on expressive lexicon size in 40 French-speaking children tested longitudinally from 22 to 48 months. The factors include those based on the lexical and phonological properties of words in the children's lexicons (phonetic complexity, word length, neighborhood density [ND], and word frequency [WF]) as well as variables measuring phonological production (percent consonants correct and phonetic inventory size). Specifically, we investigate the relative influence of these factors at individual ages, namely, 22, 29, 36, and 48 months, and which factors measured at 22 and 29 months influence lexicon size at 36 and 48 months.

Method: Children were selected based on parent-reported vocabulary size. We included children with low, medium, and high vocabulary scores. The children's lexicons were coded in terms of phonetic complexity, word length, ND, and WF, and their phonological production skills were based on measures of percent consonants correct and phonetic inventory size extracted from spontaneous speech samples at 29, 36, and 48 months. In the case of ND and WF, we focused on one- and two-syllable nouns.

Results: Across the age range, the most important factor that explained variance in lexicon size was the WF of nouns. Children who selected low-frequency nouns had larger vocabularies across all ages (22–48 months). The WF of two-syllable nouns and phonological production measured at 29 months influenced lexicon size at 36 months, whereas the WF (of one- and two-syllable words) influenced lexicon size at 48 months.

Conclusions: The findings support the role of WF and phonological production in explaining expressive vocabulary development. Children enlarge their vocabularies by adding nouns of increasingly lower frequency. Phonological production plays a role in accounting for vocabulary size up until the age of 36 months.

Supplemental Material: https://doi.org/10.23641/asha.12291074

Recent research indicates that lexical factors such as word frequency (WF) and phonological factors such as neighborhood density (ND) play an important role in determining expressive lexicon size in young children around the age of 2 years (Kern & dos Santos, 2017; Stokes, 2010, 2014; Stokes, Bleses, et al., 2012; Stokes, Kern, & dos Santos, 2012; Storkel, 2004, 2009). Little research, however, has examined the influence of these variables over time or focused on later ages, for example, 3 or 4 years of age. Yet, we know that phonological and lexical representations change over time as children acquire more words, and the processes that influence word learning change with experience (Gershkoff-Stowe, 2002; Storkel, 2009), thus emphasizing the need for studies of a longitudinal nature to examine phonological and vocabulary growth together (Edwards et al., 2011). The purpose of this study is to examine the influence of lexical and phonological variables on children's vocabulary size at four time points: 22, 29, 36, and 48 months. We extend previous studies on lexical–phonological interactions by also examining the role of the child's phonological production on vocabulary size (Kehoe et al., 2018). In particular, we would like to know whether the influence of phonological and lexical variables changes across time and whether the lexical and phonological variables, which influence children's vocabulary size around the age of 2 years (22 and 29 months), account for later vocabulary size (at 36 and 48 months). In the following sections, we outline the lexical and phonological variables employed in this study. We then review studies that have employed these variables in longitudinal research designs.

[a]Faculty of Psychology and Educational Sciences, Department of Psycholinguistics, University of Geneva, Switzerland
[b]Department of Psychology, San Diego State University, CA

Correspondence to Margaret M. Kehoe: Margaret.Winkler-Kehoe@unige.ch

Editor-in-Chief: Sean M. Redmond
Editor: Lisa M. D. Archibald

Received June 13, 2019
Revision received November 22, 2019
Accepted February 12, 2020
https://doi.org/10.1044/2020_JSLHR-19-00011

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Lexical and Phonological Variables

In this study, we distinguish between target-word variables, which are measures based on the lexical and phonological properties of words in children's lexicons, and production variables, which are based on phonological behavioral measures. Four target-word variables are measured in this study: phonetic complexity, word length, ND, and WF.[1] Two production variables are measured, namely, percent consonants correct (PCC) and syllable-initial phonetic inventory size, which are extracted from spontaneous language samples.

Target-Word Variables

Phonetic complexity[2] *and word length*. Charles Ferguson was one of the first to refer to "phonologically determined selectivity in word acquisition and use" (Ferguson & Farwell, 1975, p. 434). Since this time, a multitude of studies converge on the fact that the sound properties of words influence lexical acquisition. Evidence includes observational studies, which show that children have preferences for particular sounds (Ferguson & Farwell, 1975; Stoel-Gammon & Cooper, 1984); experimental studies, which show that children learn nonwords, which contain sounds that they can produce, more easily than nonwords, which contain sounds they cannot produce (Schwartz & Leonard, 1982); and studies based on the phonological characteristics of children's lexicons, with the latter revealing that phonological features that are acquired later in production are less frequent in young children's vocabularies (Fletcher et al., 2004; Gayraud & Kern, 2007; Stoel-Gammon, 1998). All of this evidence suggests that the phonetic complexity of the target word plays a role in vocabulary development.

Several authors have found that word length influences word learning (Jones & Brandt, 2019; Maekawa & Storkel, 2006; Storkel, 2004). Children learn short before long words. This variable takes into consideration the number of phonemes in the target word and could also be considered a measure of phonetic complexity. In this study, we examine whether phonetic complexity accounts for a similar amount of variance in word learning as word length, or whether it is even more sensitive, given the inclusion of a phonological feature's metric for each target word.

ND. ND refers to the degree of phonological similarity between a given word and a set of other words. A phonological neighbor is a word that differs from another word by substitution, deletion, or addition of a sound in any word position (Luce & Pisoni, 1998). Words that contain many phonological neighbors belong to dense neighborhoods, whereas those that contain few neighbors belong to sparse neighborhoods. In acquisition studies, ND has been shown to influence the production of a given word and the acquisition of expressive vocabulary (Coady & Aslin, 2003; German & Newman, 2004; Storkel, 2004). In a series of studies by Stokes and colleagues, ND was found to account for a high proportion of variance in the vocabulary size of children acquiring English, French, and Danish (Stokes, 2010, 2014; Stokes, Bleses, et al., 2012; Stokes, Kern, & dos Santos, 2012). In all of their studies, they coded the mean ND of one-syllable words appearing in 2-year-old children's lexicons, based on the MacArthur–Bates Communicative Development Inventories (MCDI; Fenson et al., 1993).

One general finding is that children learn words with many neighbors at earlier ages than words with fewer neighbors (Coady & Aslin, 2003; Stokes, 2010; Storkel, 2004). Stokes, Kern, and dos Santos (2012) posit that words from dense neighborhoods are less taxing on auditory–verbal short-term memories than words from sparse neighborhoods. By virtue of the fact that they share segments (e.g., balle shares neighbors with bol, bulle, belle, bas, baffe, bague, mal, salle, etc.), they offer a familiar phonetic stream that facilitates word learning.

WF. WF refers to the frequency of occurrence of words (i.e., the number of times the word has been produced). Goodman et al. (2008) investigated whether higher WF was associated with earlier word learning. They found that the relationship between frequency and acquisition was not straightforward. Higher frequency of parental input was correlated with later age of acquisition when the entire lexicon was examined. This finding reflects the fact that closed-class items, adjectives, and verbs are some of the most frequent lexical categories yet are late acquired. Within a lexical category, however, high frequency was related to earlier age of acquisition: High-frequency nouns were acquired before low-frequency ones. The authors found that the relationship between age of acquisition of a word and its frequency varied greatly according to lexical class.

Stokes and colleagues, as well as investigating the role of ND in accounting for the size of children's vocabularies, also examined the role of WF (Stokes, 2010; Stokes, Bleses, et al., 2012; Stokes, Kern, & dos Santos, 2012). Like ND, they coded the mean WF of one-syllable words. They found that WF accounted for only a small amount of vocabulary size compared to ND. More recently, Kern and dos Santos (2017) have reanalyzed Stokes, Kern, and dos Santos (2012) French data and found WF to play a stronger role. They separated out one-syllable words according to grammatical category, nouns versus predicates. Kern and dos Santos (2017) found a similar pattern of results to Stokes, Kern, and dos Santos, when nouns and predicates were grouped together, namely, a strong role for ND and a weak role for WF. However, 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. Hansen (2017) also found that frequency had a larger

[1]Phonotactic probability was not examined as a target-word measure because we analyzed real words (in contrast to phonologically controlled nonwords), which would have made its calculation complicated. As well, several studies have not found it to add unique variance beyond other lexical and phonological variables (Zamuner & Thiessen, 2018).

[2]We use the term "phonetic complexity" as coined by Jakielski (2000); however, the complexity parameters relate to phonological complexity as well.

influence on lexical development than ND in a study based on the Norwegian MCDI norms when grammatical class was controlled for.

In the current study, we introduce two innovations. First, we test whether calculating ND and WF separately for nouns (vs. predicates) offers a different perspective on the relative importance of ND and WF in explaining vocabulary size. Second, following Kehoe et al. (2018), we widen out the analyses to include one- and two-syllable words rather than limiting the analyses to one-syllable words, which reduces the generalizability of the findings.

Phonological Production Variables

By phonological production, we refer to a child's phonological capacities in spontaneous speech as measured by PCC or phonetic inventory size. There is a large body of literature that links children's vocabulary size to phonological production abilities. Children designated as late talkers (i.e., 2-year-olds who fall at or below the 10th percentile in vocabulary size) have smaller syllable structure and phonetic inventories, lower PCC and percent vowels correct, and greater numbers of phonological processes than children with average-sized vocabularies (Paul & Jennings, 1992; Rescorla & Ratner, 1996). Children with exceptionally large lexicons (i.e., 2-year-olds who are above the 85th to 90th percentile in vocabulary size) demonstrate advanced phonological abilities relative to children with average-sized lexicons (Kehoe et al., 2015, 2018; Smith et al., 2006). The association between vocabulary size and phonological production has been established across a wide variety of languages, including Italian, Cantonese, and Cypriot Greek (Fletcher et al., 2004; Petinou & Okalidou, 2006; Viterbori et al., 2018); however, few studies have examined the amount of variance in vocabulary size accounted for by phonological production versus other phonological variables (e.g., phonetic complexity and ND).

Recently, Kehoe et al. (2018) incorporated target-word and production variables into their study of factors, which influence lexicon size in 40 French-speaking children, aged 29 months. They found that three factors accounted for 76% variance in vocabulary size. The bulk of the variance was accounted for by the ND of one-syllable words (57%). In addition, the phonetic complexity of words in the children's lexicons contributed 11% unique variance and the number of syllable-initial consonants in phonetic inventories, another 8% variance. Although the amount of variance accounted for by phonological production was modest, the results, nevertheless, indicate that children who have fewer sounds in their phonetic inventories are likely to have smaller vocabularies.

Similarly, Zamuner and Thiessen (2018) found that English-speaking 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 those of Kehoe et al. (2018).

The Influence of Phonological and Lexical Variables on Lexicon Size in Longitudinal Studies

As mentioned above, only a few studies have examined the impact of lexical and phonological variables on vocabulary acquisition over time. We consider them in this section. Maekawa and Storkel (2006) observed individual differences in factors that predict age of first production of words in children (n = 3), aged 1;4 to 3;1 (years;months), which they interpreted as reflecting development effects. Word length was a consistent predictive factor across all the children, and they posited that this was a constraint operative at the earliest stages of word development. The next factor that predicted age of acquisition was phonotactic probability. It influenced word learning in the child with the youngest developmental level. WF and ND influenced word learning in the child with the most advanced stage of development. Since WF predicted acquisition for a wider range of words than ND, the authors posited that the child was using WF as a consistent cue for word acquisition, whereas he was just starting to use ND as a cue. The overall pattern was that word length and phonotactic probability preceded WF, which in turn preceded ND as factors that influence expressive vocabulary development.

Whereas Maekawa and Storkel (2006) inferred development effects from examining individual differences across age, Storkel (2009) studied developmental changes directly using data based on MCDI norms, namely, the percentage of children reported to know a word at 1-month age intervals between 16 and 30 months. She investigated the role played by phonological (i.e., phonotactic probability), lexical (i.e., ND and word length),[3] and semantic variables for this period. We focus on Storkel's (2009) findings with phonological and lexical variables. She found that phonological variables exerted a constant influence from 16 to 30 months, whereas lexical variables exerted an influence from 16 to 20 months but then diminished after that. That is, the facilitatory effect of having many lexical connections was strong through to 20 months, but adding more representations after this did not yield benefits, possibly because a certain threshold had been reached. WF was not found to have a significant influence on word learning in her study.

Comparing Maekawa and Storkel (2006) and Storkel (2009), we observe some commonalities and differences. Both studies suggest that the relative influence of target-word variables such as ND, WF, and word length changes over time. Maekawa and Storkel propose, however, that ND plays a role later on, whereas Storkel (2009) found that ND plays a role early on. Maekawa and Storkel found that WF predicted word acquisition, whereas Storkel (2009) did not. Neither of these studies tested children beyond the age of 3 years, nor did they integrate target-word and production variables into one study, hence the importance of this study.

[3]In contrast to Storkel (2009), we refer to ND and word length as phonological rather than lexical variables since they refer to the phonological form of the word.

Apart from studying the relative influence of phonological and lexical variables on vocabulary growth across time, we are also interested in whether variables measured at earlier time points explain variance at later time points. Stokes, Kern, and dos Santos, (2012) hypothesized that late talkers, who select words with high NDs, may have more difficulty proceeding to later stages of word learning than children who select words with low NDs. They may have got stuck in a period of "extended statistical learning," unable to map words from sparser neighborhoods to semantic representations. This suggests that the ND of children's lexicons at 2 years of age may explain later vocabulary scores. We intend to investigate whether ND and/or other phonological and lexical variables measured around the age of 2 years account for vocabulary development at 3 and 4 years of age.

Current Study and Research Questions

The current study is an extension of Kehoe et al. (2018), which examined the influence of target-word and production variables on the lexicon size of 40 French-speaking children, aged 29 months. We extend Kehoe et al. (2018) by separating the lexicon into grammatical class, nouns and predicates (although we focus only on nouns in the current study; see below), and by introducing the variables WF and word length. We add a longitudinal perspective by studying the children at an earlier (22 months) and two later (36 and 48 months) time points. We use a vocabulary checklist designed for children aged 36 and 48 months, Developpement du Langage de Production en Français (DLPF) Versions 3 and 4 (Bassano et al., 2005), to extract the target-word properties of the lexicon at the older ages (36 and 48 months) and the European-French version of the MCDI, l'Inventaire Français du Développement Communicatif (IFDC; Kern & Gayraud, 2010), to extract target-word variables at the younger ages (22 and 29 months). We collect phonological production measures (PCC and phonetic inventory size) at 29, 36, and 48 months. Due to the original design of the study, the focus of which was not lexical–phonology connections, we do not have phonological production measures at 22 months. Please note that, after 30 to 36 months, it is generally acknowledged that it is an impossible task to capture all lexical items that children are using (Feldman et al., 2005). We nevertheless assume that parent report at 36 and 48 months will provide a relative estimate of the word types children are using, although it may not provide an exact estimate of the number of types.

The overarching goal of our study is to investigate the influence of lexical and phonological factors on vocabulary growth in a group of French-speaking children. The specific research aims are as follows: First, we examine what factors among target-word and phonological production variables best account for vocabulary size at individual time points (target-word variables at 22 months and target-word and phonological production variables at 29, 36, and 48 months). Stokes and colleagues' studies would suggest that ND accounts for the greatest amount of variance at ages 22 and 29 months (Stokes, 2010, 2014; Stokes, Bleses, et al., 2012; Stokes, Kern, & dos Santos, 2012), although separation of the lexicon according to grammatical class with a focus on nouns may yield more favorable results for WF (Hansen, 2017; Kern & dos Santos, 2017). Kehoe et al.'s (2018) findings would also suggest that phonological production plays a role at 29 months. Our predictions for the analysis of the lexicon at 36 and 48 months are less clear since few authors have examined target-word effects at later ages, although experimental word-learning studies would suggest that they still play a role (Storkel & Lee, 2011). Nevertheless, it is likely that measures based on phonological production do not account for additional variance in vocabulary size since children's consonantal acquisition is largely complete by 3 and 4 years of age. Thus, we predict a strong role for ND, possibly in combination with WF and phonological production at the younger ages, and a diminishing role for phonological production at the later ages.

Second, we investigate which lexical and phonological factors measured at 22 and 29 months best account for expressive vocabulary size at 36 and 48 months. The findings of Stokes and colleagues would lead us to believe that ND plays the major role in explaining vocabulary size (Stokes, Kern, & dos Santos, 2012). In both research questions, expressive vocabulary size is determined by parental report.

Method

This study is part of a larger project, the purpose of which was to examine the association between early language comprehension and later language and literacy development. The larger study involved testing 65 French-speaking children longitudinally from the ages of 16 months to 5;0.

Participants

Children were recruited from birth lists provided by the state of Geneva to the university. They were tested within a 2- to 3-week interval of the age points: 16, 22, 29, 36, 48, and 60 months. In this study, we focus on the ages 22, 29, 36, and 48 months.[4] We selected 40 children from the larger group (n = 65) based on their vocabulary scores at 29 months. We did not include the entire sample due to the time-consuming nature of the analyses. We chose 29 months (rather than an earlier time point) because previous analyses showed it to be a more stable age for explaining later language development on the basis of expressive vocabulary (Patrucco-Nanchen et al., 2019). We then examined the results of the children at the earlier time point (22 months) and the two later time points (36 and 48 months). In order to have children with a diverse range of vocabulary scores, we employed the selection criteria of Kehoe et al. (2018): Four groups were selected: (a) Late 1 (n = 8; three girls) obtained IFDC scores at or below the 10th percentile (range: 40 to 221 words), (b) Late 2 (n = 9; six girls)

[4]We did not analyze the children at 16 months because there were too much missing data. At 60 months, there was no information on lexicon size or phonological production.

obtained scores between the 15th and 25th percentile (range: 268 to 353), (c) Middle (n = 11; five girls) obtained scores between the 40th and 60th percentile (range: 372 to 474), and (d) Precocious (n = 12; six girls) obtained scores that exceeded the 90th percentile (range: 572 to 677). All children were monolingual, had normal hearing, were reported to be in good health, and were developing normally. The children came from middle to upper-middle class backgrounds. Previous analyses of the data showed no influence of socioeconomic status on vocabulary development (Friend et al., 2019). All parents of children included in the study completed an informed consent form, which was approved by the University of Geneva's ethics committee.

General Procedure

At each visit, children attended a single session of 60 min 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. The executive function and morphosyntactic tests were included in the larger study, but they are not reported in this study. At 29, 36, and 48 months, the children engaged in a play session while interacting with one of their parents. In addition, the parents completed the IFDC at or prior to the session when the children were aged 22 and 29 months. They completed the DLPF Versions 3 and 4 when the children were aged 36 and 48 months, respectively.

Vocabulary Measures

Parent Questionnaires

IFDC. Parents completed the *Mots et Phrases* form of the IFDC (Kern & Gayraud, 2010). This form is designed for children from 16 to 30 months. It consists of a list of 690 words organized in 22 semantic categories. Parents are asked to indicate whether their child produces the word. The IFDC is sensitive to vocabulary development over time and has strong short-term test–retest reliability (r = .90; Kern & Gayraud, 2010). At 29 months of age, the sample-specific internal consistency of the measure was high (α = .98).

DLPF Versions 3 and 4. DLPF (Bassano et al., 2005) is a parental questionnaire designed to test the language development of children aged from 2 years through to the end of their fourth year. We focus on the vocabulary component of the test. The authors drew on longitudinal corpora of French-speaking children and their own experimental studies to compile the vocabulary items and categories. In our study, we utilized Versions 3 and 4, which are designed to test children aged 3;0 to 4;0.[5] Both versions contain 23 semantic categories of words divided into four grammatical categories (onomatopoeia and games, nouns, predicates, and grammatical words), making a total of 1,233 and 1,488 words, respectively. Parents are asked to indicate whether their child produces the word. The DLPF is currently being normed on 517 children, aged 18 to 42 months; thus, we are unable to provide psychometric data on this test. The sample-specific internal consistency of the measure was high (at 36 months: α = .99; at 48 months: α = .99).

Spontaneous Language Sample

The spontaneous language sample was used to extract the phonological production measures, PCC and phonetic inventory size. It involved the recording of one parent and their child playing with an extended Fisher Price farm play set. The play items were the same for each child, thus ensuring a uniform set of vocabulary items per child. The toys were also selected to elicit a wide variety of phonemes. The duration of the language samples was 20 min at 29 months, 15 min at 36 months, and 10 min at 48 months. The language samples were recorded using a portable digital tape recorder (Marantz PMD620). The parent was instructed to play with the child just as they would do at home.

Phonological production. The phonological production measures were obtained using Phon, a software program specifically designed for the analysis of phonological data (Rose et al., 2006). The digitalized recordings of the language samples were segmented into utterances, glossed, and phonetically transcribed. A team of French-speaking undergraduate and graduate students (n = 10), who had experience in phonetic transcription and had taken courses in articulatory and acoustic phonetics, performed the analyses. Two measures of phonological production were obtained: PCC and the number of consonants in syllable-initial position in the children's phonetic inventories (referred to as PhonInv in the tables and appendixes). A consonant was designated as being part of the phonetic inventory if it was present at least 2 times and in two different words. We also obtained a third measure, namely, the number of consonants in syllable-final position. However, this variable was not found to contribute to any of the statistical analyses and, thus, will not be considered further. Calculations of PCC were computed automatically for each child based on the entire number of utterances in the language sample using the query function PCC-PVC in Phon. The PCC calculations were based on an average number of 308 words (range: 77 to 712) at 29 months, 414 words (93 to 808) at 36 months, and 338 words (75 to 652) at 48 months. Three participants at each age range were retranscribed by a second transcriber using the Blind Transcription function of Phon. Point-to-point agreement in terms of consonant transcription was good: .97 at 29 months, .89 at 36 months, and .96 at 48 months.

Data Coding of the Target-Word Measures

The coding of target-word variables was based on reduced versions of the IFDC and the DLPF3 and DLPF4.

[5]The DLPF4 is actually designed for children, aged 3;1–3;6 and older, whereas we used it with children aged 4;0. We did not observe any ceiling effects with this group of children (M = 1283, range: 592–1,472; the number of vocabulary items on the DLPF4 is 1,488).

Following the procedures of Stokes and colleagues (Stokes, 2010; Stokes, Kern, & dos Santos, 2012), the reduced set of the IFDC included 12 categories of items considered representative of core vocabulary. It omitted onomatopoeia, games and routines, and context-based items (e.g., people and function words). As with the IFDC, onomatopoeia and context-based items were excluded from the DLPF3 and DLPF4, resulting in 15 categories of items. The total number of items was 518 for the IFDC, 895 for the DLPF3, and 1,093 for the DLPF4. The data were then coded for grammatical class, noun versus predicate, and number of syllables, one versus two syllables.

Phonetic Complexity

Using the Index of Phonetic Complexity (Jakielski, 2000), a point was assigned to each word of the reduced versions of the IFDC, DLPF3, and DLPF4 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 (e.g., balle [**b**a**l**] "ball," which has labial and coronal places of articulation); a tautosyllabic cluster, a cluster that occurs within a syllable (e.g., crayon [**kʁ**ejɔ̃] "pencil"); or a heterosyllabic cluster, a cluster that is split across two syllables (e.g., tracteur [tʁa**kt**œːʁ] "tractor"). The Index of Phonetic Complexity also assigns points to rhotic vowels, but since rhotic vowels do not occur in French, this category was excluded.

Word Length

All words in the reduced versions of the IFDC, DLPF3, and DLPF4 were coded for word length in phonemes. French does not have falling diphthongs, so all vowel nuclei were counted as one phoneme. French has rising diphthongs, made up of a glide and a vowel (e.g., toi [twa] "you"). The glide was counted as a phoneme.

ND

One- and two-syllable nouns of the reduced versions of the IFDC, DLPF3, and DLPF4 were coded for ND using the values generated by the Lexique3 database, a corpus of adult language (New et al., 2007). The most frequent phonological forms were chosen when two noun choices were provided (e.g., **figure**/visage "face").

WF

One- and two-syllable nouns of the reduced versions of the IFDC, DLPF3, and DLPF4 were coded for token frequency using the Lexique3 database (New et al., 2007). The frequency data are based on film subtitles from French films and French translations of English films and television series. In the case of nouns that can have multiple inflections, the frequency value corresponded to the sum of all possible inflections that had the same pronunciation as the target form (the most frequent phonological form).

Some readers may question why we have used adult rather than child corpora to calculate ND and WF values. We follow the arguments of Stokes and colleagues (Stokes, 2014; Stokes, Kern, & dos Santos, 2012) in noting that (a) children are exposed not only to child-directed but also to adult-directed speech, (b) ND and WF information from child corpora is not readily available in French, and (c) authors have observed high correlations between WF values generated in child- and adult-directed corpora (Gierut & Dale, 2007; Stokes, 2010).

Once coding was completed, mean phonetic complexity and word length were determined for all words in the children's reduced lexicons (IFDC, DLPF3, DLPF4). Mean ND values were obtained for one- and two-syllable nouns. WF values were first log transformed due to the skewed nature of the raw frequency values (Brysbaert et al., 2018), and then mean values were calculated for one- and two-syllable nouns. When calculations of ND and WF were based on fewer than four items (in the category of one- and two-syllable nouns), these data were removed, and consequently, certain children were excluded from statistical analyses at a given time point (see Results section for the exact number of children excluded in each analysis).

Although we obtained ND and WF values for predicates as well, we present the findings based on nouns for three reasons. First, the inclusion of predicates would have necessitated taking other factors into consideration such as imageability (verbs are less imageable than nouns), which was beyond the scope of the current study. Second, the inclusion of variables for predicates resulted in an unwieldy number of variables given the present sample size. Finally, Kern and dos Santos (2017) did not find that ND and WF based on predicates influenced lexicon size at least at 24 to 30 months.

Statistical Analyses

The statistical analyses were performed using R statistical software (R Core Team, 2016). The amount of variance in expressive vocabulary size explained by the predictor variables was determined by linear regression (lm function in R). We used the total lexicon size based on parent report as our dependent variable. Preliminary analyses revealed that similar correlations were obtained regardless of whether we took the total versus reduced lexicon size as the dependent variable.

Target-word predictor variables were phonetic complexity, word length, ND of one-syllable nouns (ND1N), ND of two-syllable nouns (ND2N), log-transformed WF of one-syllable nouns (LWF1N), and log-transformed WF of two-syllable nouns (LWF2N). Production predictor variables were the two variables extracted from spontaneous speech: PCC and syllable-initial consonant inventory size. We used the approach of backward elimination to establish the best statistical model. All variables were entered, and then nonsignificant variables were removed in a step-by-step fashion. Once we obtained a model in which only significant predictors remained, we entered each variable separately to determine the amount of individual variance accounted for by each predictor. The only exception we made was in the case of phonetic complexity and word length, which were highly correlated with each other. We ran two separate models: one with phonetic complexity and one with word length (see Zamuner & Thiessen, 2018, for a similar approach with other variables). In some cases, word length emerged as significant (at 29 months), but not phonetic complexity, and in other cases, phonetic complexity emerged as significant (at 36 and 48 months), but not word length. A priori power analyses indicated that sample sizes were sufficient to detect small effects at power = .95 (G*Power: Faul et al., 2007).

Results

Descriptive Statistics

The Appendix displays the means, standard deviations, and ranges for lexicon size, as well as the target-word variables and production variables across the four different ages (22, 29, 36, and 48 months). Figures S1–S9 in Supplemental Material S1 present the results graphically across age for the 40 children. In all the analyses, children from the four vocabulary levels are collapsed into a single group.

Accounting for Vocabulary Size Across Age

The first research question examines what factors account for vocabulary size at each time point.

Correlation Matrices

Tables S1–S4 of Supplemental Material S2 display the correlations between vocabulary size ($IFDC_{tot}$) and the target-word and production variables, as well as display the correlations among the variables at 22 to 48 months. At 22 months, moderate to moderately high significant correlations were observed between vocabulary size and all variables with the exception of the ND and WF of two-syllable nouns (see Table S1 of Supplemental Material S2). As mentioned, no production measures were obtained at this age. At 29 months, moderate to high significant correlations were observed between vocabulary size and all variables. There were also many significant correlations among the variables (see Table S2 of Supplemental Material S2). At 36 months, moderate to high significant correlations were observed between vocabulary size and all variables. However, in comparison to 29 months, correlations for ND and production variables have declined in strength, and correlations for the WF of two-syllable words have increased (see Table S3 of Supplemental Material S2). At 48 months, WF and phonetic complexity (as well as word length) are highly correlated with vocabulary size; ND and the production variables are no longer significantly correlated with it (see Table S4 of Supplemental Material S2).

Accounting for Vocabulary Size at 22 Months

To conduct our statistical analyses at 22 months, we had to eliminate several children: Three children (one Late 1, one Late 2, and one Middle) did not produce sufficient numbers of one- and two-syllable nouns, and an additional two children (two Late 2) did not produce sufficient numbers of one-syllable nouns (although they did produce two-syllable nouns). Using a database of 35 children, we obtained a significant model, $F(2, 32) = 15.25$, $p < .001$, with one variable, the WF of one-syllable words, accounting for 42% variance in vocabulary size. In addition, the ND of one-syllable words was marginally significant ($t = -1.85$, $p = .07$), accounting for 4% of variance. The results of this model are presented in Table 1.

Accounting for Vocabulary Size at 29 Months

To conduct our analyses at 29 months, one child (Late 1) was eliminated because he did not produce sufficient numbers of one-syllable nouns. We obtained a significant model, $F(5, 33) = 67.57$, $p < .001$, with five variables influencing vocabulary size and accounting for 90% of variance: The WF of one-syllable nouns accounted for the highest unique variance (79%), followed by word length (an additional 5% variance).[6] The remaining three variables (WF of two-syllable nouns, phonetic inventory size, and ND of one-syllable nouns) accounted for 5% to 6% variance in lexicon size. The results of this model are presented in Table 2.

Accounting for Vocabulary Size at 36 Months

The database at 36 months consisted of 39 children since one child (Precocious) had not been administered the DLPF3. Multiple regression indicated a significant model, $F(3, 35) = 68.26$, $p < .001$, with two variables influencing vocabulary size and accounting for 83% of variance (see Table 3). The WF of one-syllable nouns accounted for the highest unique variance, 69%, and phonetic complexity, an additional 14% unique variance. As well, phonetic inventory size was marginally significant and contributed 1% variance ($t = 1.9$, $p = .065$), resulting in a total of 84% variance accounted for.

Accounting for Vocabulary Size at 48 Months

Multiple regression analyses conducted when the children were 48 months of age indicated a significant model, $F(2, 37) = 135.6$, $p < .001$, with two variables influencing vocabulary size and accounting for 87% of variance (see Table 4). The WF of two-syllable words accounted for the highest variance, 86%, and phonetic complexity, an additional 1% unique variance. It must be noted, however, that there were high correlations between the WF of one- and two-syllable nouns and between WF and phonetic complexity (see Table S4 of Supplemental Material S2; VIFs in excess of 4.0), meaning that it is difficult to isolate the individual effects of these three variables (see Zamuner & Thiessen, 2018).

[6]LWF1N and WL were highly correlated (.79), but the variance inflation factor was acceptable (2.7).

Table 1. Results of multiple regression analyses for predicting vocabulary size at 22 months based on lexical and phonological variables measured at 22 months.

Model	β	*SE*	*t*	*p* value	$R^{2\Delta}$
$IFDC_{tot}$ ~ LWF1NX + (ND1N; *n* = 35)					
Intercept	1131.600	173.739	6.513	< .001	
LWF1N	–370.539	125.969	–2.942	.006	.416
ND1N	–11.039	5.974	–1.848	.074	.456

Note. LWF1N = log-transformed word frequency of one-syllable nouns; ND1N = neighborhood density of one-syllable nouns.

Composition of Children's Vocabularies: One Versus Two Syllables

The above analyses indicated that the WF of one-syllable words was more strongly correlated with lexicon size at the younger age ranges, but over time, the WF of two-syllable words came to be more strongly correlated with it. One reason for this could be the composition of children's vocabularies, in which one-syllable words predominate at the younger ages and two-syllable words predominate at the older ages.

We examined the proportion of one- and two-syllable nouns (i.e., the number of one- or two-syllable nouns as a proportion of the total number of one- and two-syllable nouns) in children's lexicons at each age range. The findings are shown in Figure 1. Children's lexicons contained more two- than one-syllable words at each age range, although the gap between the proportion of one- versus two-syllable words widened with age: at 22 months, there was a .05 difference between the proportions of one- versus two-syllable words (i.e., .47 vs. .52), whereas at 48 months, there was a .12 difference (i.e., .44 vs. .56).

Summary

Multiple regression analyses across the different age ranges indicated that a high proportion of variance in vocabulary size could be accounted for by WF. The WF of one-syllable words played the strongest role at 22, 29, and 36 months; and the WF of two-syllable words, at 48 months. In addition, small degrees of variance were accounted for by the ND of one-syllable nouns (at 22 and 29 months), word length (at 29 months), and phonetic complexity (at 36 and 48 months). The production variable, phonetic inventory size, also added unique variance at 29 months and was marginally significant at 36 months.

Table 2. Results of multiple regression analyses for predicting vocabulary size at 29 months based on lexical and phonological variables measured at 29 months.

Model	β	*SE*	*t*	*p* value	$R^{2\Delta}$
$IFDC_{tot}$ ~ LWF1N + WL + LWF2N + PhonInv + ND1N (*n* = 39)					
Intercept	1835.981	659.763	2.783	.009	
LWF1N	–681.940	201.891	–3.378	.002	.792
WL	226.359	79.269	2.856	.007	.845
LWF2N	–752.264	217.195	–3.464	.002	.865
PhonInv	11.793	5.172	2.280	.03	.887
ND1N	–21.660	10.353	–2.092	.04	.898

Note. LWF1N = log-transformed word frequency of one-syllable nouns; WL = word length; LWF2N = log-transformed word frequency of two-syllable nouns; PhonInv = syllable-initial phonetic inventory size; ND1N = neighborhood density of one-syllable nouns.

Accounting for Vocabulary Size at 36 and 48 Months

The second research question examined which variables at 22 and 29 months account for vocabulary size at 36 and 48 months. Table S5 of Supplemental Material S2 summarizes the correlations between predictor variables at 22 and 29 months and outcome variables at 36 and 48 months. In fact, there were no significant correlations between variables at 22 months and vocabulary size at 36 and 48 months. Therefore, no variables at 22 months were entered into regression analyses. However, all predictor variables at 29 months were correlated with vocabulary size at 36 months, and several of them were correlated with vocabulary size at 48 months; the exceptions were ND of two-syllable nouns and phonetic inventory size. All correlated variables at 29 months were entered into regression analyses.

Three variables measured at 29 months influenced vocabulary size at 36 months, $F(3, 35) = 26.78$, $p < .001$. They were the WF of two-syllable nouns, PCC, and phonetic inventory size. Together, they accounted for 67% of variance in vocabulary size. Phonetic inventory size accounted for 41% variance; the WF of two-syllable nouns accounted for 20% variance, and PCC accounted for the remaining 6% variance.

Two variables at 29 months accounted for vocabulary size at 48 months, $F(2, 36) = 14.36$, $p < .001$, explaining 41% of the variance. They were the WF of two-syllable nouns, which accounted for 35% variance, and the WF of one-syllable nouns, which accounted for the remaining 6% variance. Table 5 summarizes the findings of the regression analyses for accounting for vocabulary size at ages 36 and 48 months based on variables measured at 29 months.

Summary

Regression analyses indicated that production variables in combination with WF at 29 months accounted for vocabulary size at 36 months, whereas WF on its own explained vocabulary size at 48 months. In both cases, the WF of two-syllable words better accounted for vocabulary size than the WF of one-syllable words.

Discussion

The aim of this study was to examine the influence of phonological and lexical factors on vocabulary size in a group of French-speaking children. It extends the work of previous authors by following the children longitudinally from 22 to 48 months and by including information on

Table 3. Results of multiple regression analyses for predicting vocabulary size at 36 months based on lexical and phonological variables measured at 36 months.

Model	β	*SE*	*t*	*p* value	$R^{2\Delta}$
$DLPF3_{tot}$ ~ LWF1N + PhonCom + (PhonInv)					
Intercept	3201.845	730.031	4.386	< .001	
LWF1N	−2933.066	304.168	−9.643	< .001	.687
PhonCom	541.704	117.121	4.625	< .001	.830
PhonInv	18.164	9.555	1.901	.065	.842

Note. LWF1N = log-transformed word frequency of one-syllable nouns; PhonCom = phonetic complexity; PhonInv = syllable-initial phonetic inventory size.

children's phonological production alongside information on the phonological and lexical characteristics of their lexicons. We also separated the lexicon according to grammatical class (nouns vs. predicates), focusing on nouns.

Our results show that, across age, a high degree of variance in vocabulary size can be explained by the WF of nouns. We also found that the WF of two-syllable nouns and phonological production play a role in accounting for lexicon size at later age ranges. In the following paragraphs, we summarize the results in more detail and discuss their significance for understanding the role of frequency and phonology in lexical acquisition.

Accounting for Vocabulary Size Over Time

The first research question examined what factors account for vocabulary size over time. Previous research has given a contradictory picture of the relative importance of lexical and phonological factors in explaining vocabulary development longitudinally. For example, Maekawa and Storkel's (2006) study suggested that ND plays a role in word learning in children, aged up to 36 months, whereas Storkel's (2009) study found little effect of ND beyond 20 months. Neither of these studies looked at children older than 36 months of age. More recently, studies by Stokes and colleagues indicate that ND is the key factor in explaining vocabulary size in children aged 24 to 30 months.

Thus, our prediction was that ND would play the strongest role in explaining lexicon size at the youngest ages (22 and 29 months). We did not make firm predictions for the older ages, given the lack of research. We tempered our predictions for the younger ages, nevertheless, given recent work by Kern and dos Santos (2017), which showed that WF played a stronger role than ND in explaining variance in vocabulary size once grammatical category was taken into consideration. Consequently, we also calculated ND and WF separately for nouns and predicates, focusing on nouns in the current study. Our results are consistent with those of Kern and dos Santos (2017) in showing that the WF of nouns plays the key role in accounting for lexicon size. This was the case not only at 22 and 29 months but also at the older ages as well. In the following paragraphs, we comment on the role of the different sets of variables: WF versus ND, phonetic complexity versus word length, and phonological production.

Table 4. Results of multiple regression analyses for predicting vocabulary size at 48 months based on lexical and phonological variables measured at 48 months.

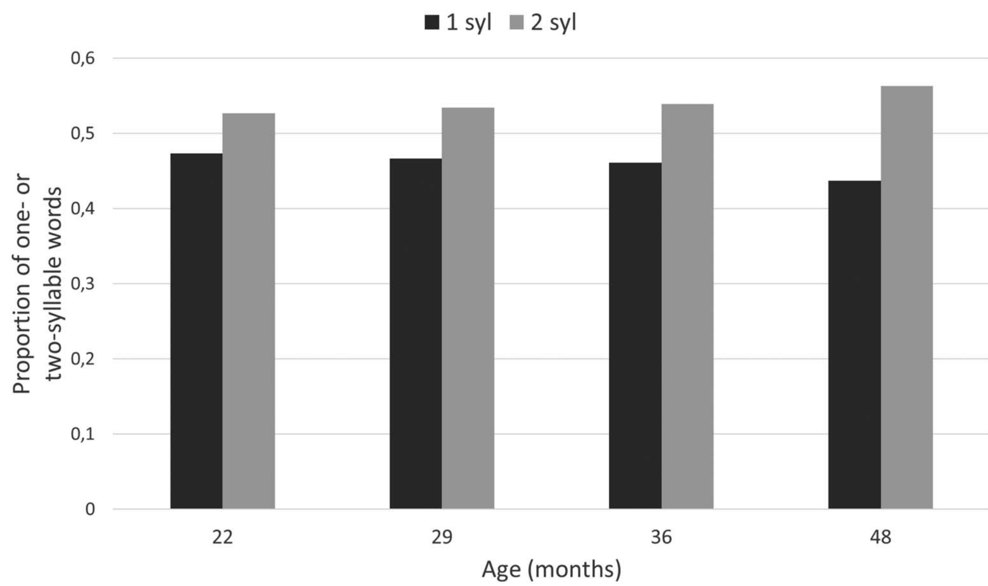
Model	β	*SE*	*t*	*p* value	$R^{2\Delta}$
$DLPF4_{tot}$ ~ LWF2N + PhonCom					
Intercept	580.6	2193.3	0.265	.793	
LWF2N	−2588.7	531.8	−4.868	< .001	.861
PhonCom	923.2	421.7	2.189	.035	.873

Note. LWF2N = log-transformed word frequency of two-syllable nouns; PhonCom = phonetic complexity.

WF Versus ND

The findings indicate a high negative correlation between WF and lexicon size: Children who have larger-sized lexicons select nouns with lower frequency. Our findings, thus, support Kern and dos Santos (2017) and Hansen (2017) in showing that, when grammatical class is taken into consideration, frequency plays a more important role than ND in explaining vocabulary size at least with respect to nouns. Of course, one may wonder which procedure is the most valid for determining lexical and phonological effects on lexicon size: one that takes grammatical class into consideration or one that does not. Like Goodman et al. (2008), we do not assume that children are able to analyze their lexicons into nouns and predicates; however, we rely on their evidence that "the effect of frequency on vocabulary acquisition interacts with semantic-syntactic categories" (p. 527). From a statistical point of view, the effects of WF are minimized when grammatical class is not taken into consideration due to the large frequency differences between nouns and predicates. Furthermore, we assume that, if ND plays a stronger role than WF in vocabulary development, its effect should be present even when controlling for grammatical category. Our results show that ND declines in importance with advancing age, whereas WF gains in importance. This is evident from the correlation matrices, in which the correlation coefficients between the ND of one-syllable words and lexicon size decrease (in absolute magnitude) from −.77 at 29 months to .15 at 48 months, whereas the correlation coefficients between the WF of one-syllable words and lexicon size stay high across

Figure 1. Proportions of one- and two-syllable (syl) nouns in children's lexicons across age.



all ages (−.66 to −.89), and the ones for WF of two-syllable words increase (−.26 to −.93; see Tables S1–S4 of Supplemental Material S2 and also Figure S4 of the Supplemental Material S1, which show few developmental changes in ND1N beyond 29 months in contrast to WF, in which changes are apparent—see Figures S6 and S7 of Supplemental Material S1).

Phonetic Complexity vs. Word Length

In this study, we included two variables that tap the phonetic complexity of the target word: one that takes phonological features into account and one that refers to the number of phonemes in the word. One question we posed was whether they account for similar degrees of variance in vocabulary size. The overall answer appears to be "yes." As can be observed from correlation matrices across time (see Tables S1–S4 in Supplemental Material S2), phonetic complexity and word length were highly correlated (correlation coefficients from .86 to .93). Nevertheless, there were subtle differences in how they influenced lexicon size over time. Word length accounted for 5% unique variance at 29 months. Phonetic complexity accounted for 14% unique variance in lexicon size at 36 months and a tiny proportion at 48 months (i.e., 1%). Thus, an interpretation of these findings could be that, at younger ages, children are sensitive to the number of phonemes in a word when adding vocabulary items, whereas, at older ages, children are sensitive to the finer phonetic features of a word. We may wonder why children should be sensitive to word length since this variable does not correspond to any linguistic structural unit. For example, few phonological processes pay attention to the number of phonemes in a word. We hypothesize that this variable is indirectly tapping children's increasing ability to process words with complex syllable structure, that is, words containing codas and clusters.

Table 5. Results of multiple regression analyses for predicting vocabulary size at 36 and 48 months on the basis of variables measured at 29 months.

Model	β	*SE*	*t*	*p* value	$R^{2\Delta}$
Predicting vocabulary size at 36 months					
Model: $DLPF3_{tot} \sim C_{SI}$ + LWF2N + PCC (*n* = 39)					
Intercept	2294.500	610.319	3.760	< .001	
PhonInv	0.626	0.142	4.415	< .001	.407
LWF2N	−2006.845	430.079	−4.666	< .001	.613
PCC	6.924	2.554	2.711	.01	.671
Predicting vocabulary size at 48 months					
Model: $DLPF4_{tot} \sim$ LWF2N + LWF1N (*n* = 39)					
Intercept	4492.9	628.9	7.144	< .001	
LWF2N	−1387.7	580.1	−2.392	.02	.350
LWF1N	−850.2	347.2	−2.449	.02	.413

Note. Only variables at 29 months were entered into the models because variables at 22 months were not significantly correlated with vocabulary size at 36 and 48 months. PhonInv = syllable-initial phonetic inventory size; LWF2N = log-transformed word frequency of two-syllable nouns; PCC = percent consonants correct; LWF1N = log-transformed word frequency of one-syllable nouns.

Production Variables

The correlation matrices (see Tables S1–S4 in Supplemental Material S2) show that production variables correlate quite strongly with vocabulary size at 29 months (correlation coefficients from .62 to .72) and to a (low–)moderate degree at 36 months (.35 to .43), but they no longer correlate with vocabulary size at 48 months. We hypothesize that they would have been strongly correlated with vocabulary size at 22 months, but we do not have the data to confirm this. In our regression analyses, phonetic inventory size was the only production variable that accounted for unique variance. It accounted for 2% unique variance at 29 months and had a marginal role at 36 months. Overall, it seems that production variables when placed alongside target-word variables do not account for much variance in vocabulary size at individual age ranges. These results differ from those of Kehoe et al. (2018) who found that phonetic inventory size accounted for a larger proportion of variance (i.e., 8%) at 29 months; however, this study did not take WF into consideration. Once WF was entered into the regression models, the effect of phonological production largely disappeared.

Accounting for Vocabulary Size at 36 and 48 Months

The second research question asked what factors measured at earlier time points (22 and 29 months) account for vocabulary size at later time points (36 and 48 months). We found that no variable measured at 22 months was correlated with vocabulary size at later ages. This finding is consistent with diverse reports that indicate that the age range under 2 years is a more precarious one for explaining later language outcomes than the age range above 2 years (Paul, 1996; Rescorla, 2013).

We found that several factors at 29 months accounted for vocabulary size at 36 months. In particular, phonological production played an important part. Children who had larger numbers of consonants in their syllable-initial phonetic inventories and who had better phonological production scores had larger lexicons at 36 months. Even at this relatively advanced point in phonological development (average PCC is 80% at 29 months), having well-developed phonological skills aids lexical learning. This may be for the reasons outlined by researchers who have explored lexical–phonology connections (McAllister Byun & Tessier, 2016; Vihman, 2014, 2017; Zamuner & Thiessen, 2018): Words that contain sounds that children can produce are more salient to them, are more easily processed, place fewer demands on phonological memory, and, thus, are more likely to be learned.

Apart from phonology, the WF of two-syllable words accounted for unique variance (approximately 20%) at 36 months. It also accounted for the greatest degree of unique variance at 48 months (i.e., 35%). The WF of one-syllable words accounted for the greatest degree of variance at individual time points (at least through to 36 months); however, the WF of two-syllable words was the one most strongly related to later vocabulary size. This may come about for several reasons. As Figure 1 shows, French children's lexicons from 22 to 48 months contain more two- than one-syllable words, and the relative difference between one- and two-syllable words increases with age. The WF of two-syllable words may have played a stronger role in explaining later vocabulary size than one-syllable words due to the higher proportions of two-syllable words in the lexicons of children. In this respect, the study of a language that contains a different composition of one- and two-syllable words to that of French would be illuminating.

The importance of WF for two-syllable words in explaining later vocabulary size may also relate to their increased complexity. Overall, two-syllable nouns are more complex than one-syllable nouns. They are characterized by greater phonetic complexity and length. They have fewer phonological neighbors and are less frequent. If children are able to acquire more complex low-frequency words at an earlier time point, presumably, they have better word learning strategies, which will then allow them to acquire a greater number of words at a later time point (see as well Figures S6 vs. S7 of Supplemental Material S1, which show stronger developmental effects for the WF of two-syllable as compared to one-syllable nouns).

The Role of Frequency and Phonology in Vocabulary Development

The main finding of this study was that the WF of nouns played the greatest role in accounting for variance in vocabulary size over time. The importance of noun frequency in vocabulary acquisition is somewhat paradoxical since, in terms of overall vocabulary, nouns are less frequent than other grammatical classes. Hence, to fully understand this effect, we would need to consider factors responsible for the noun bias in vocabulary development (Gentner, 1982; Tardif et al., 1999), a topic that extends beyond the realms of this article.

Our findings reinforce what has been said before, namely, that the frequency at which a child is exposed to a given word influences the ultimate learning of this word. As Hoff and Naigles (2002) expound, apart from the number of repeated exposures of a word that act as multiple learning trials, words presented frequently are likely to be presented in diverse contexts and syntactic frames, leading to cross-situational learning and syntactic bootstrapping. In an exemplar-based approach, the greater number of times a child hears and produces a word, the more robust will be his or her lexical representations, which in turn leads to higher level phonological knowledge and more successful word learning (Edwards et al., 2011; Pierrehumbert, 2003).

Storkel (2009) points out that the variable WF reflects the number of times a person encounters the form of a word and its meaning; thus, it can be considered to incorporate both phonological and semantic aspects of words. In her study, WF did not account for unique variance in age-of-acquisition of words over time, whereas phonological (phonological probability), lexical (ND, word length), and semantic (semantic density) variables did. She posited that WF may not have had a significant influence on word learning in her study because it was subsumed by other (phonological, lexical, and semantic) variables. WF accounted for a high proportion of variance in vocabulary size in the current study even when placed alongside phonological and lexical variables. It remains to be seen, however, whether the inclusion of variables that tap semantic knowledge may have reduced the influence of WF. An indication in this direction comes from a recent study by Jones and Brandt (2019) who found that WF played a modest role in comparison to the semantic variable of concreteness in explaining expressive vocabulary development.

At the moment, our results tell us that children who learn words of low frequency will have larger lexicon sizes at a given time point and will have larger lexicons at later stages of development, but they do not tell us why some children learn low-frequency words easily and some do not. To answer this question, we may need to look beyond phonological and lexical influences. Studies that have examined individual differences in learning rate in fast mapping and word learning experiments (Gray, 2005; Gray & Brinkley, 2011), or that have examined the influence of children's speech processing skills (Fernald et al., 2006; Hurtado et al., 2008; Song et al., 2018) or more general processing and cognitive abilities (Kan & Windsor, 2010) on vocabulary development, may hold clues as to why some children learn low-frequency words more easily than others.

A focus of the study was to determine whether phonological production influences vocabulary development over and beyond the effects of target-word variables. We observed the influence of phonological production through to 36 months but no longer at 48 months. Children who had better expressive phonology skills at 29 months had larger lexicons at 29 and 36 months. Many studies suggest that a potential word candidate is more salient to a child if it contains sounds that the child has produced or can produce. Keren-Portnoy et al. (2010) showed that phonological strings that have been articulated are represented more robustly in memory than strings that have been heard but not articulated. Zamuner and Thiessen (2018) found that children were more likely to imitate words that contain sounds that they have previously produced accurately. Vihman (2017) summarizes the results of several experiments that support her "articulatory filter" hypothesis, namely, that the child's own level of speech sound production influences the way he or she represents and processes (non)words containing these sounds. All these findings indicate that phonological production is highly relevant to word learning.

Limitations of the Study

One of the main limitations of this study relates to the procedure itself, which examines the effect of a reduced set of variables on vocabulary development. We aimed to determine how far we could take phonological and lexical variables in explaining lexical development but, as such, ignored the role of semantic, social–pragmatic, demographic, and behavioral variables, which need to be factored in to arrive at a complete understanding of variables influencing vocabulary development. Another danger of this type of methodology is the high degree of multicollinearity between phonological and lexical variables (Storkel, 2004; Zamuner & Thiessen, 2018). We relied on statistical models to delineate which factors played the most important role, but they may not reflect the way these variables play out in the mind of the child. All of these variables may influence vocabulary development at any one time. In addition, defining the variables of WF and ND according to word length in syllables or grammatical class may be artificial delimitations for the child.

As noted, children were selected for the study based on their IFDC scores at 29 months. This selection procedure had the advantage of ensuring a great deal of variability in children's lexical skills, but it may have led to inflated correlations. Lower correlations might have been obtained with a different participant selection procedure. Smolik (2019) points out that studies that have examined the relation between target-word variables (e.g., ND) and vocabulary size have yielded stronger correlations than those using age-of-acquisition of individual words as the dependent variable. Thus, weaker effects may have been obtained if we had used a different dependent variable. Other concerns include the fact that we have no psychometric results yet on the DLPF; thus, caution should be employed in the interpretation of the parent report results at 36 and 48 months. It would also have been worthwhile including a single-word production task in addition to the spontaneous speech sample to have a more complete understanding of the children's phonological production abilities.

In this study, we ignored the role of predicates in explaining vocabulary size. Even though variables specific to nouns ended up accounting for high degrees of variance in lexicon size, opening up the study to include predicates would be important (see Kern & dos Santos, 2016). More recently, findings suggest inclusion of the variable imageability might allow for an analysis of the role of lexical and phonological effects on word learning across diverse grammatical categories (Hansen, 2017; Smolik, 2019). Finally, future studies should validate the results obtained on ND and WF for adult-directed corpora with child-directed ones.

Conclusion

An important contribution of this study is the finding that children from the age of 22 to 48 months enlarge their vocabularies by adding nouns of increasingly lower frequency. Our findings do not support previous studies that point to ND as the determining factor in explaining vocabulary size. We found that phonological production (but not ND) plays a role in accounting for vocabulary size up until the age of 36 months. This finding is consistent with the idea that children are more likely to find salient, and therefore to learn, new words that follow phonological patterns in which they are already proficient. In combination, these findings suggest that children progressively add lower frequency words to their vocabularies with the constraint that words conform to their phonological production capacities.

Acknowledgments

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. We would like to thank Marilyn Vihman, Christophe dos

Santos, and Sophie Kern for their advice and helpful discussion. We would also like to thank the following students for their assistance in phonetic transcription and data coding: Lucile da Silva, Elisa Chaplin, Pauline Mudry, Mélanie Anzalone, Nolwen Beraud, Samantha Cirieco, Julie Franco, Valérie Giuseppina Rezzonico, Mathilde Jolicoeur, and Ombeline Laurie Vaglio-Agnes.

References

Bassano, D., Labrell, F., Champaud, C., Lemétayer, F., & Bonnet, P. (2005). Le DLPF: Un nouvel outil pour l'évaluation du Developpement du Langage de Production en Français [The DLPF: A new tool for assessing the development of language production in French]. *Enfance, 57,* 171–208. https://www.cairn.info/revue-enfance1-2005-2-page-171.htm

Brysbaert, M., Madera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Coady, J., & Aslin, R. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language, 30*(2), 441–469. https://doi.org/10.1017/S0305000903005579

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*(1), 35–40. https://doi.org/10.1017/S0305000910000450

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feldman, H., Campbell, T., Kurs-Lasky, M., Rockette, H., Dale, P., Colborn, D. K., & Paradise, J. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856–868. https://doi.org/10.1111/j.1467-8624.2005.00882.x

Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *MacArthur Communicative Development Inventories: User's guide and technical manual*. Singular.

Ferguson, C., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*(2), 419–439. https://doi.org/10.2307/412864

Fernald, A., Perfors, A., & Marchman, V. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

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*(6–8), 535–545. https://doi.org/10.1080/02699200410001703655

Friend, M., Smolak, E., Patrucco-Nanchen, T., & Zesiger, P. (2019). Language status at age 3: Group and individual prediction from vocabulary comprehension in the second year. *Developmental Psychology, 55*(1), 9–22. https://doi.org/10.1037/dev0000617

Gayraud, F., & Kern, S. (2007). Caractéristiques phonologiques des noms en fonction de l'âge d'acquisition [Phonological characteristics of nouns as a function of age of acquisition]. *Enfance, 59,* 324–338. https://www.cairn.info/revue-enfance1-2007-4-page-324.htm

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Volume 2. Language, thought and culture* (pp. 301–334). Erlbaum.

German, D., & Newman, R. (2004). The impact of lexical factors on children's word finding errors. *Journal of Speech, Language, and Hearing Research, 47*(3), 624–636. https://doi.org/10.1044/1092-4388(2004/048)

Gershkoff-Stowe, L. (2002). Object-naming, vocabulary growth and the development of word retrieval abilities. *Journal of Memory and Language, 46*(4), 665–687. https://doi.org/10.1006/jmla.2001.2830

Gierut, J. A., & Dale, R. A. (2007). Comparability of lexical corpora: Word frequency in phonological generalization. *Clinical Linguistics & Phonetics, 21*(6), 423–433. https://doi.org/10.1080/02699200701299891

Goodman, J., Dale, P., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary? *Journal of Child Language, 35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research, 48*(6), 1452–1467. https://doi.org/10.1044/1092-4388(2005/101)

Gray, S., & Brinkley, S. (2011). Fast mapping and word learning by preschoolers with SLI in a supported learning context: Effect of encoding cues, phonotactic probability and object familiarity. *Journal of Speech, Language, and Hearing Research, 54*(3), 870–884. https://doi.org/10.1044/1092-4388(2010/09-0285)

Hansen, P. (2017). What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological density on lexical development. *First Language, 37*(2), 205–225. https://doi.org/10.1177/0142723716679956

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

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

Jones, S., & Brandt, S. (2019). Neighborhood density and word production in delayed and advanced learners. *Journal of Speech, Language, and Hearing Research, 62*(8), 2847–2854. https://doi.org/10.1044/2019_JSLHR-L-18-0468

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kehoe, M., Chaplin, E., Mudry, P., & Friend, M. (2015). La relation entre le développement du lexique et de la phonologie chez les enfants francophones [The development of phonological skills in late and early talkers]. *Rééducation Orthophonique, 263,* 61–85.

Kehoe, M., Patrucco-Nanchen, T., Friend, M., & Zesiger, P. (2018). The relation between phonological and lexical development in French-speaking children. *Clinical Linguistics & Phonetics, 32*(12), 1103–1125. https://doi.org/10.1080/02699206.2018.1510984

Keren-Portnoy, T., Vihman, M., DePaolis, R., Whitaker, C., & Williams, N. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research, 53*(3), 1280–1293. https://doi.org/10.1044/1092-4388(2009/09-0003)

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

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). John Benjamins. https://doi.org/10.1075/tilar.22.10ker

Kern, S., & Gayraud, F. (2010). *l'Inventaire Français du Développement Communicatif* [The French Inventory of Communicative Development]. Editions La Cigale.

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. https://doi.org/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*(3), 439–459. https://doi.org/10.1017/S0305000906007458

McAllister Byun, T., & Tessier, A.-M. (2016). Motor influences in grammar in an emergentist model of phonology. *Language and Linguistics Compass, 10*(9), 431–452. https://doi.org/10.1111/lnc3.12205

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*(4), 661–677. https://doi.org/10.1017/S014271640707035X

Patrucco-Nanchen, T., Friend, M., Poulin-Dubois, D., & Zesiger, P. (2019). Which infants' lexical skills predict language skills at 3 years? A longitudinal study. *Infant Behavior & Development, 57,* 101379. https://doi.org/10.1016/j.infbeh.2019.101379

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5*(2), 5–21. https://doi.org/10.1044/1058-0360.0502.05

Paul, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. *Journal of Speech and Hearing Research, 35*(1), 99–107. https://doi.org/10.1044/jshr.3501.99

Petinou, K., & Okalidou, A. (2006). Speech patterns in Cypriot-Greek late talkers. *Applied Psycholinguistics, 27*(3), 335–353. https://doi.org/10.1017/S0142716406060309

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*(2–3), 115–154. https://doi.org/10.1177/00238309030460020501

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

Rescorla, L. (2013). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. https://doi.org/10.1002/ddrr.1108

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research, 39*(1), 153–165. https://doi.org/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). Cascadilla Press.

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*(2), 319–336. https://doi.org/10.1017/S0305000900004748

Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics, 27*(3), 355–375. https://doi.org/10.1017/S0142716406060310

Smolik, F. (2019). Imageability and neighborhood density facilitates the age of word acquisition in Czech. *Journal of Speech, Language, and Hearing Research, 62*(5), 1403–1415. https://doi.org/10.1044/2018_JSLHR-L-18-0242

Song, J., Demuth, K., & Morgan, J. (2018). Input and processing factors affecting infants' vocabulary size at 19 and 25 months. *Frontiers of Psychology, 9,* Article No. 2398. https://doi.org/10.3389/fpsyg.2018.02398

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). Brookes.

Stoel-Gammon, C., & Cooper, J. (1984). Patterns of early lexical and phonological development. *Journal of Child Language, 11*(2), 247–271. https://doi.org/10.1017/S0305000900005766

Stokes, S. (2010). Neighborhood density and word frequency in toddlers. *Journal of Speech, Language, and Hearing Research, 53*(3), 670–683. https://doi.org/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*(3), 634–657. https://doi.org/10.1017/S030500091300010X

Stokes, S., Bleses, D., Basbøll, H., & Lambertsen, C. (2012). Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech and Hearing Research, 55*(5), 1265–1273. https://doi.org/10.1044/1092-4388(2012/10-0291)

Stokes, S., Kern, S., & dos Santos, C. D. (2012). Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language, 39*(1), 105–129. https://doi.org/10.1017/S0305000911000031

Storkel, H. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25*(2), 201–221. https://doi.org/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*(2), 291–321. https://doi.org/10.1017/S030500090800891X

Storkel, H., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*(2), 191–211. https://doi.org/10.1080/01690961003787609

Tardif, T., Gelman, S., & Xu, F. (1999). Putting the 'noun bias' in context: A comparison of English and Mandarin. *Journal of Child Language, 24*(3), 535–565. https://doi.org/10.1017/S030500099700319X

Vihman, M. (2014). *Phonological development: The first two years.* Wiley-Blackwell.

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

Viterbori, P., Zanobini, M., & Cozzani, F. (2018). Phonological development in children with different lexical skills. *First Language, 38*(5), 538–559. https://doi.org/10.1177/0142723718784369

Zamuner, T., & Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics, 63*(4), 609–632. https://doi.org/10.1017/cnj.2018.10

Appendix

Descriptive Statistics for Lexicon Size, Target-Word, and Production Variables at the Four Different Age Points

Variables	*n*	*M* (*SD*)	Minimum	Maximum
22 months				
$IFDC_{tot}$	37	192.35 (143.40)	26	523
PhonCom	37	3.25 (0.63)	1.33	3.99
WL	37	4.19 (0.46)	2.63	4.94
ND1N	35	26.00 (3.84)	19.31	36.71
ND2N	34	8.88 (1.32)	5.82	11.75
LWF1N	35	1.74 (0.18)	1.52	2.28
LWF2N	34	1.37 (0.13)	1.04	1.67
29 months				
$IFDC_{tot}$	40	402.28 (183.19)	40	677
PhonCom	40	3.65 (0.26)	2.83	4.07
WL	40	4.44 (0.21)	3.94	4.80
ND1N	39	23.80 (1.40)	22.21	27.33
ND2N	40	8.59 (0.63)	7.53	10.40
LWF1N	39	1.62 (0.09)	1.49	1.88
LWF2N	40	1.32 (0.05)	1.23	1.44
PCC	40	79.94 (10.10)	42.81	95.45
PhonInv	40	15.65 (2.73)	6	19
36 months				
$DLPF3_{tot}$	39	892.67 (241.39)	360	1233
PhonCom	39	3.70 (0.16)	2.88	3.83
WL	39	4.31 (0.12)	3.67	4.41
ND1N	39	23.64 (0.73)	22.63	26.91
ND2N	39	8.45 (0.48)	7.88	10.94
LWF1N	39	1.58 (0.06)	1.52	1.71
LWF2N	39	1.21 (0.05)	1.14	1.35
PCC	39	86.66 (8.23)	61.15	97.23
PhonInv	39	17.36 (1.81)	12	19
48 months				
$DLPF4_{tot}$	40	1283.75 (199.35)	592	1472
PhonCom	40	3.89 (0.06)	3.65	3.97
WL	40	4.42 (0.04)	4.33	4.48
ND1N	40	23.75 (1.37)	22.43	29.64
ND2N	40	8.30 (0.86)	7.65	12.44
LWF1N	40	1.49 (0.05)	1.44	1.70
LWF2N	40	1.12 (0.05)	1.05	1.29
PCC	40	95.43 (3.00)	87.34	99.61
PhonInv	40	17.63 (1.13)	14	19

Note. $IFDC_{tot}$ = total vocabulary score on the l'Inventaire Français du Développement Communicatif (the French version of the MCDI); PhonCom = phonetic complexity; WL = word length; ND1N = neighborhood density of one-syllable nouns; ND2N = neighborhood density of two-syllable nouns; LWF1N = log-transformed word frequency of one-syllable nouns; LWF2N = log-transformed word frequency of two-syllable nouns; PCC = percent consonants correct; PhonInv = syllable-initial phonetic inventory; $DLPF3_{tot}$ = total vocabulary score on the Developpement du Langage de Production en Français Version 3; $DLPF4_{tot}$ = total vocabulary score on the Developpement du Langage de Production en Français Version 4.