



HAL
open science

Invariance in variation: Frequency and neighbourhood density as predictors of vocabulary size

Sophie Kern, Christophe dos Santos

► **To cite this version:**

Sophie Kern, Christophe dos Santos. Invariance in variation: Frequency and neighbourhood density as predictors of vocabulary size. Maya Hickmann; Edy Veneziano; Harriet Jisa. Sources of variation in first language acquisition: Languages, contexts, and learners, 22, John Benjamins, pp.183-200, 2018, Trends in Language Acquisition Research, 9789027244123. 10.1075/tilar.22.10ker . hal-01996411

HAL Id: hal-01996411

<https://hal.science/hal-01996411v1>

Submitted on 5 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

This author-accepted manuscript is under copyright.

Sophie Kern, Christophe dos Santos. Chapter 9. Invariance in variation: frequency and neighbourhood density as predictors of vocabulary size. *Sources of variation in first language acquisition: Languages, contexts, and learners*, 22, pp.183-200, 2018, Trends in Language Acquisition Research, (<https://benjamins.com/catalog/tilar.22.10ker>) (<https://doi.org/10.1075/tilar.22.10ker>)

Invariance in variation: frequency and neighbourhood density as predictors of vocabulary size

S. Kern (Laboratoire Dynamique Du Langage) & C. dos Santos (Université de Tours)

Abstract

This article examines the influence of word frequency (WF) and neighbourhood density (ND) in vocabulary acquisition of French children. Data were collected through the French version of the MacArthur Development Inventory. A regression analysis based on 462 children aged between 16 and 30 months who have acquired at least 5 words revealed that ND and WF together predicted 45% of the variance in vocabulary size, with ND and WF uniquely accounting for 32.2% and 12.8% of that variance respectively. The same analysis was done with nouns and predicates only. For nouns, the model predicted 64.6% of the variance whereas for predicates, the size of predicate vocabulary was not correlated with any of the two variables.

Key words: lexicon, development, frequency, neighbourhood density, children, French

Introduction

Child's early productive lexicon is quantitatively and qualitatively very different from the adult language (Fenson, Dale, Reznick, Bates, Hartung, Pethick & Reilly, 1993). Several factors play a role on the quantitative and qualitative development of first word production. Among them, two factors were frequently taken into consideration: phonetic/phonological development (Vihman, 1996) and input characteristics (Lieven, 2010). In this paper, we will examine more deeply the influence of input and in particular the influence of frequency and neighbourhood density (ND) on children's early lexical development. Facilitative effects of high density and frequency were demonstrated on language processing, on speech recognition as well as on speech production across the lifespan (Ellis, 2002; Vitevitch, 2002; Vitevitch & Sommers, 2003). The questions we will answer are the followings: does frequency and/or ND play a role on lexical development? And do they have the same role as a function of grammatical categories?

Theoretical background

The role of input is considered as important to early language learning within a function-based perspective (*e.g.* Cameron-Faulkner, Lieven & Tomasello, 2003; Gallaway & Richards, 1994; Snow, 1977). Even if most of the research conducted to date provides only indirect forms of evidence for the evaluation of the effect of input on vocabulary acquisition, there also appears to be a general theoretical consensus on the positive effect of input on word learning. There are many ways to characterize language input, in a qualitative manner following typological descriptions or in a quantitative manner, looking at the frequency of different units, or even by combining both ways. In this paper, we will focus on two words characteristics: frequency of occurrence in the input and ND.

Frequency & lexical acquisition

A large amount of studies agrees on the role of word frequency (WF) on lexical development in children acquiring their mother tongue. Important correlations between the lexicon size of children and the amount of heard input were observed. (Hart & Risley, 1995; Weizman & Snow, 2001). Furthermore, typically developing children as well as specific language impaired (SLI) children acquire more easily words they have been frequently exposed to (Rice, Oetting, Marquis, Bode & Pae, 1994). In addition, frequency of exposure to a specific grammatical category seems to help the learning of this specific category (Gopnik & Choi, 1990, 1995). Goodman, Dale & Li (2008) came to the conclusion that even if it is true that frequency of exposure to a word plays an important role in the acquisition of the word, this is only a part of the truth. The authors have correlated the age of acquisition evaluated through the use of the MacArthur-Bates Communicative Inventories (Fenson *et al.*, 1993) of 562 words in English-learning children aged from 8 to 30 months with the frequency of these words in 28 corpora of child-directed speech available on the Childes website (<http://childes.psy.cmu.edu>). Interestingly, they showed on the one hand, that if all words were considered, there was a positive correlation between age of acquisition and frequency: the frequent words were acquired later than the less frequent ones. They explained this result by a late acquisition by children of closed class items. On the other hand, inside the lexical categories of verbs and nouns, a high frequency was associated with an early acquisition. Furthermore, frequency played a more important role in the productive vocabulary than in the receptive one: in comprehension, parental frequency was correlated to language acquisition only for common nouns.

According to all the presented studies, a link between frequency of exposure and language development exists. However, given the limited number of studies and the disparities in the findings and in the measurements of WF, it is difficult to make a clear statement regarding the

role of WF and measures of production. In addition, up to now and to our knowledge, no one except Goodman *et al.* (2008), was able to explain why particular words were acquired before others inside one specific grammatical category.

Neighbourhood density and lexical acquisition

The second variable we are going to take into consideration for vocabulary learning is ND, that is to say the link between word acquisition order and how many phonological neighbours one word has. Two words are considered as phonological neighbours as soon as they are composed of the same phonemes with the exception of one (Charles-Luce & Luce, 1990). The difference between one word and its phonological neighbours can be due to a phoneme substitution, the addition of one phoneme or the deletion of one phoneme. In a specific language, words have more or less phonological neighbours. For example, the word ‘balle’ (ball) [bal] has 42 neighbours whereas the word ‘fenêtre ’ (window) [fənɛtʁ] has only one. Words with a lot of phonological neighbours belong to a dense neighbourhood whereas words with few neighbours belong to a sparse neighbourhood.

Several studies have investigated the role of ND on lexical development in young children. Charles-Luce and Luce (1990, 1995) as well as Logan (1992) demonstrated that children’s words have fewer neighbours than the same words in the adult’s lexicon. By considering absolute numbers of neighbourhoods, they also showed a trend toward denser neighbourhoods with age. However, these studies were limited to either children’s expressive or receptive vocabularies, have underestimated the size of children’s vocabularies and more importantly they haven’t normalized neighbourhoods by the sizes of the vocabularies considered.

Coady & Aslin (2003) conducted three more comprehensive analyses of phonological neighbourhoods, trying to take the above mentioned limitations into account. Phonological neighbourhoods were calculated for all monosyllabic words produced by two English-speaking children from the age of 2;3 until the age of 3;6 and their mothers. In addition, ND

was calculated for an adult lexicon. The results support previous findings with ND being denser in the adult lexicon than those for the same words in the developing lexicon. Data suggests as well that children are acquiring words from denser than average neighbourhoods: for words that appear in children's lexicons, the average number of phonological neighbours in the adult lexicon is higher than the average for all of the words in the adult lexicon. Finally, they show that words in the developing lexicons had roughly twice as many neighbours than did previous analyses (6.5 neighbours vs. 2.25-3.32 neighbours). This difference in neighbours' numbers can be due to length differences as previous studies considered all word length. A second study by Coady and Aslin seems to confirm this idea. In a second study, Coady and Aslin dealt with the relationship between vocabulary sizes, word length and ND. They were able to show, that ND decreases as word length increases. But they also showed that children's lexicons contained a lot of shorter words, with progressively fewer neighbours as word length increased. So, because children have an important proportion of the shorter words that reside in denser phonological neighbourhoods, ND should decrease over development as children later acquire the longer words with sparser phonological neighbourhoods. Finally, in a last analysis, Coady and Aslin tried to evaluate the relationship between ND and vocabulary size by calculating the ND relative to vocabulary size in monosyllabic words only. The calculated ratios showed that in proportion to vocabulary size, ND decreased between the age of 3;6 and adulthood.

Frequency and neighbourhood density

Although ND is positively correlated with WF (Landauer & Steeter, 1973), and negatively correlated with word length (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985), only a few studies have considered their influence of early lexicon acquisition at the same time. Storkel (2004a and 2009) studied the relationship between ND, WF, word length and age of acquisition by looking at nouns produced by American-speaking children, from 8 to 30 months of age. Data

were available from a cross-sectional sample of 1800 American children. The database consists of the percentage of children from the norming sample who were reported to know each of the MCDI words at 1-month age intervals between 8 to 30 months. The results mirrored those of previously reported studies: high-density words are acquired earlier than low-density words. But new observations also emerged: the effect of ND was only evident for low-frequency words, not for high-frequency words. In addition, ND predicted the age of acquisition for short words but not for long words. This finding suggests that ND may play a lesser role when learning high-frequency or long words. Concerning the effect of WF, early acquired words were higher in frequency than later acquired words and this effect was more robust for short words than for long words. Finally, early acquired words were shorter than later acquired words, but this effect was present only for high-frequency words.

According to Storkel (2009), WF is a composite variable playing possibly a role in three different linguistic domains: phonology, lexicon and semantics. Consequently, in her model for predicting the age of word acquisition, she used two phonological predictors (for each word, she calculated the mean frequency of segments and diphones according to their position in the word) and two lexical predictors (ND and word length). In the end, she concluded to the influence of both phonological variables from 16 to 30 months and to an influence of both lexical predictors from 16 to 20 months only.

Maekawa & Storkel (2006) also attempted to differentiate the effects of ND, WF and word length on expressive vocabulary development in 3 children between the ages of 1;4 and 3;1 native speakers of American English with typical language development. Naturalistic conversational samples were obtained from CHILDES. As high phonotactic probability seems to facilitate both expressive and receptive lexical acquisition especially in children with smaller vocabularies (Edwards, Beckman & Munson, 2004; Storkel & Rogers, 2000), they added phonotactic probability to the analysis. They also kept in mind the positive correlation

between phonotactic probability and ND (Storkel, 2004b). The study identified length as a predictor of expressive vocabulary development across three subjects whereas the other three factors affected only one child each. This inter-child variability could be explained by differences in the words sampled across children but also by developmental differences across children, suggesting that the role of the factors changes across development. According to the authors, children are first constrained by word length in their lexical acquisition, before being able to use phonotactic probabilities. In a third step, they are supposed to lean on frequency of items and finally in a last step use ND to develop their expressive lexicon.

In 2010, Stokes studied the influence of frequency and ND in the lexical development of 222 British-English-learning children (mean age of 27 months). In line with previous studies, only monosyllabic content words (160 nouns, 88 verbs and 31 adjectives) were included. Due to lack of consensus surrounding the issue of the relation of ND and WF, both factors were investigated as separate variables before collinearity of the two was examined in a regression analysis. Data were collected through the British version of the MacArthur CDI developed by Klee & Harrison (2001). Stokes came to the conclusion that ND and WF were responsible of 61% of the lexicon size variance with 47% and 14% respectively. ND was inversely related to vocabulary size: as vocabulary size increased, more words from sparse neighbourhoods were added. WF was positively related to vocabulary size, with more frequent words in larger vocabularies. Moreover, she pointed out the fact that low-vocabulary children scored significantly higher on ND and significantly lower on WF than did high-vocabulary children but there was more variability in ND and WF for children at the lowest points of the vocabulary continuum. To explore the cross-linguistic validity of these conclusions, the same analyses were conducted on a French-speaking and a Danish-speaking population. The expressive lexicons of 208 French-speaking two-year-old children were coded for ND and WF (Stokes, Kern, dos Santos, 2012). Regression revealed that ND and WF together

predicted 62% of the variance in vocabulary size, with ND and WF uniquely accounting for 53% and 9% of that variance respectively. The research by Stokes, Bleses, Basbøll, Lambertsen (2012) explored the impact of ND, WF and word length (WL) on the vocabulary size of 894 Danish-speaking two-year-old children. Regression revealed that ND, WF, word length and age together predicted 47% of the variance in vocabulary size, with ND, WF, WL and Age uniquely accounting for 39%, 3.2%, 2.2% and 2.8% of that variance respectively. Children with small vocabularies had learned words that were denser, more frequent in the ambient language and shorter than the words of children with larger vocabularies. The strong role of ND in emerging languages found in other languages was replicated in Danish. However, the role of WF was much smaller than in English and French. This less important role has been explained by a different distribution of word classes on the parental checklist.

All these studies have highlighted the important role of ND and more mixed conclusions concerning the role of WF on age of acquisition. A very restricted set of languages has been considered and one could expect different results for ND due to structural differences as ND varies as a function of language (Vitevitch & Stamer, 2006). This point will be discussed in the conclusion section. Furthermore, very few studies were developmental studies on a large time-span which could be of interest to show the precise influence of ND and WF according to age and/or lexical size. Finally, the majority of the above described studies focalized on the productive lexicon and more exactly on monosyllabic content words.

In our paper, we will explore the role of ND and WF on lexical acquisition in French-learning children from 16 to 30 months of age. We will concentrate our analyses on nouns and predicates produced between 16 and 30 months. In a following paper, we intend to consider receptive lexicon as well as words longer than one syllable.

Method

Subjects

522 monolingual French-speaking children between 16 and 30 months participated to the study.

Table 1: population

FCDI	Words and sentences														
Age	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number	11	10	0	10	28	49	39	35	88	40	30	24	34	35	29
Total	462														

Data collection

To evaluate the lexical level of children, a parental report adapted and normed to French: IFDC (Kern & Gayraud, 2010) - Inventaire Français du Développement Communicatif Mots et Phrases - was documented by subjects' mothers. This parental report is the French version of the MacArthur-Bates communicative development inventory created by Elisabeth Bates and normed by Fenson et al. (1993).

This version, aimed to children between 16 and 30 months of age, is composed of two main parts. The first evaluates the productive lexicon: 690 words distributed in 22 semantic categories and four grammatical ones. The second part evaluates the morpho-syntactic level of children.

Data processing and coding

Data reduction

Following Stokes (2010) and Stokes, Kern & dos Santos (2012) methods, we restricted our analysis to monosyllabic content words (nouns, verbs, adjectives and adverbs). The selected

words were at minimum composed of one vowel and one or more consonants. Words with two vowels among which one was a schwa in a non-accentuated syllable and only preceded by one single consonant were considered monosyllabic words (for example, the word ‘cheveux’ (hair) [ʃəvø], which contains a schwa in a non-accentuated syllable, was considered monosyllabic as it is frequently produced [ʃvø] in colloquial French). Finally, concerning verbs, among several possible monosyllabic forms, the one with the highest frequency has been chosen (for example, for the verb ‘dire’ (to say), the more frequent monosyllabic form is [di]). Eventually, a list of 220 mots consisting of 131 nouns, 56 verbs, 30 adjectives and 3 adverbs was included in the analysis.

Frequency

The token frequency of each word was determined through the Lexique3 database (New, Brysbaert, Veronis, & Pallier, 2007). Lexique3 contains more than 50 million French words. Oral frequency (from film subtitles) of each word is given according to its grammatical nature.

Neighbourhood density

ND of each word was calculated based on the most frequent monosyllabic phonological form from the same grammatical category. For example, ND of the verb ‘chanter’ (to sing) was calculated based on the monosyllabic phonological form (/ʃãt/ (sing) only, which is the most frequent monosyllabic phonological form of this verb. Then, the phonological form /ʃãt/ has 30 neighbours (ND=30 ; /pãt/ ‘pente’ (slope), /sãt/ ‘sente’ (footpath), /ʃãs/ ‘chance’ (luck),...)

Results

Instead of computing a mean value for WF and ND for each child as it was the case in Stokes, Kern & dos Santos (2012), we decided to compute the median value for WF and ND for each child. This choice was made to neutralize the effect of extreme values, specifically for WF

data. Half of the words have a WF of 47.31 or less, but the mean of the whole set of words is 251.75 with the maximum being the verb ‘avoir’ (to have) with a WF of 15267.71. Moreover, we log-transformed the frequency data (see appendix for a dataset summary).

In order to have a better image of each vocabulary set and to have enough children by age group, we decided to only select the children who were able to produce at least 5 words out of the 220 words selected for this study. Then, the medians for WF and ND for each child were standardized by age group to neutralize the effect of age. A within age-group z-score was computed for all variables.

ND and WF as predictors of total vocabulary size

Our first question about the influence of WF and ND on vocabulary size concerns the vocabulary size of each child for all the 220 monosyllabic words (cf. above). A first analysis of correlations among variables shows that vocabulary size is moderately and negatively correlated with WF ($r(462) = -0.49, p < 0.01$) and ND ($r(462) = -0.57, p < 0.01$). From these correlations, we can say that smaller vocabularies consist of more frequent words than larger ones as well as words with a higher WF and ND (WF and ND are weakly and positively correlated ($r(462) = 0.25, p < 0.01$)).

A multiple regression analysis was conducted. The variable to predict was vocabulary size. WF and ND were the predictors. The model is significant ($F(2, 459) = 189.31, p < 0.01$). WF and ND account for 45% of the variance in vocabulary size. ND is the strongest predictor according to the t values (Table 2). A hierarchical multiple regression showed that ND accounts for 32.2% of the variance in vocabulary size ($F(1, 460) = 176.59, p < 0.01$) and WF for 12.8% ($F(1, 459) = 107.68, p < 0.01$).

Children who present a small vocabulary size tend to produce words with high ND and high WF. Larger vocabularies tend to be composed of more words from sparse neighbourhood and low frequency.

Table 2: Table of coefficients for the multiple regression predicting total vocabulary size

	Standardized coefficients			
	β	t	p	Confidence interval (95%)
(Constant)		0.00	1.00	
ND	-0.48	-13.29	0.00	-0.55 to -0.40
WF	-0.37	-10.38	0.00	-0.44 to -0.30

Following Goodman *et al.* (2008), the effect of frequency on vocabulary acquisition interacts with semantic-syntactic categories, we divided the 220 monosyllabic CDI words of this study into two groups. The first group contains 131 words from the category of nouns and the second group contains 89 words from the category of predicates. The predicate category is composed of verbs, adjectives and adverbs, as they all emerge after nouns in children's lexicon. As previously, we selected only children who produced at least five words in the category under study. For nouns, it included 456 children and for predicates 382 children. The question was to know if ND and WF have a different impact on vocabulary size in these two categories as Goodman *et al.* (2008) suggest for frequency.

ND and WF predictors of noun vocabulary size

To answer the question about the influence of ND and WF on noun vocabulary size, a first step was done in looking at correlation among variables. Noun vocabulary size is negatively correlated with ND and this correlation is close to a strong correlation level ($r(456) = -0.68$; $p < 0.01$). Noun vocabulary size is also negatively correlated with WF but strongly correlated ($r(456) = -0.70$; $p < 0.01$). Then, when noun vocabulary size increases, ND and WF decrease (ND and WF are positively and moderately correlated ($r(456) = 0.50$; $p < 0.01$).

The sign of this correlation is the same as previously found for total vocabulary size but the magnitude of the correlation is stronger with noun vocabulary size.

Plots of these relationships are shown in Figures 1 and 2. The two plots for noun vocabulary size by ND and WF respectively reflect the significant negative correlation, with a low number of nouns being comprised of high NDs and high IFs relative to larger vocabularies.

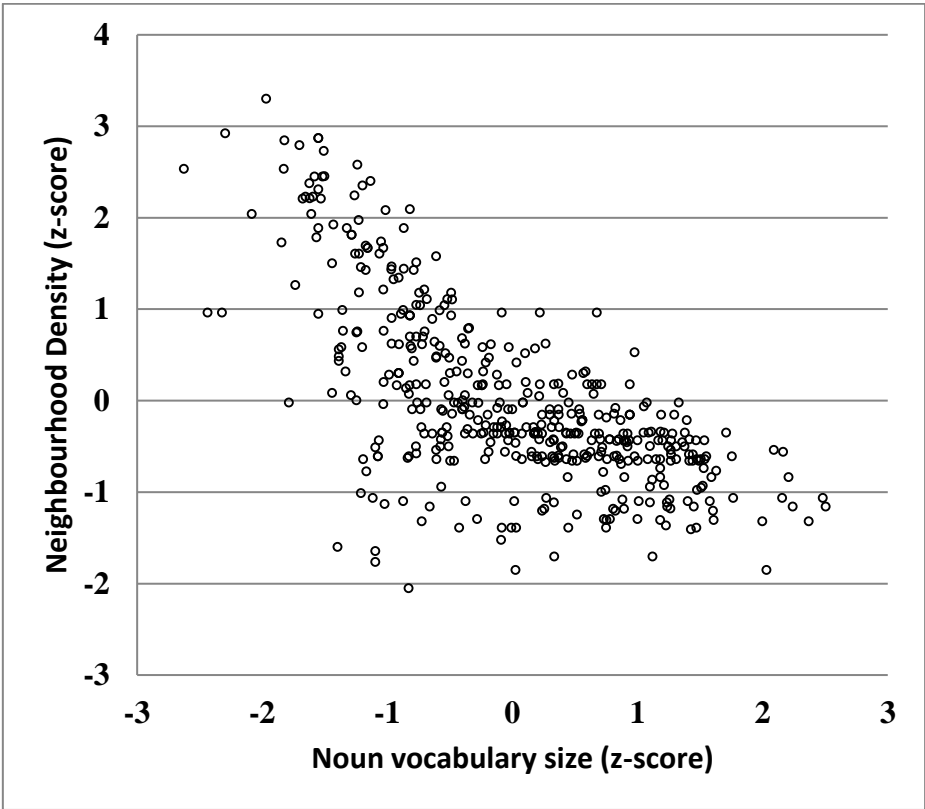


Figure 1: Scatterplot of ND by noun vocabulary size.

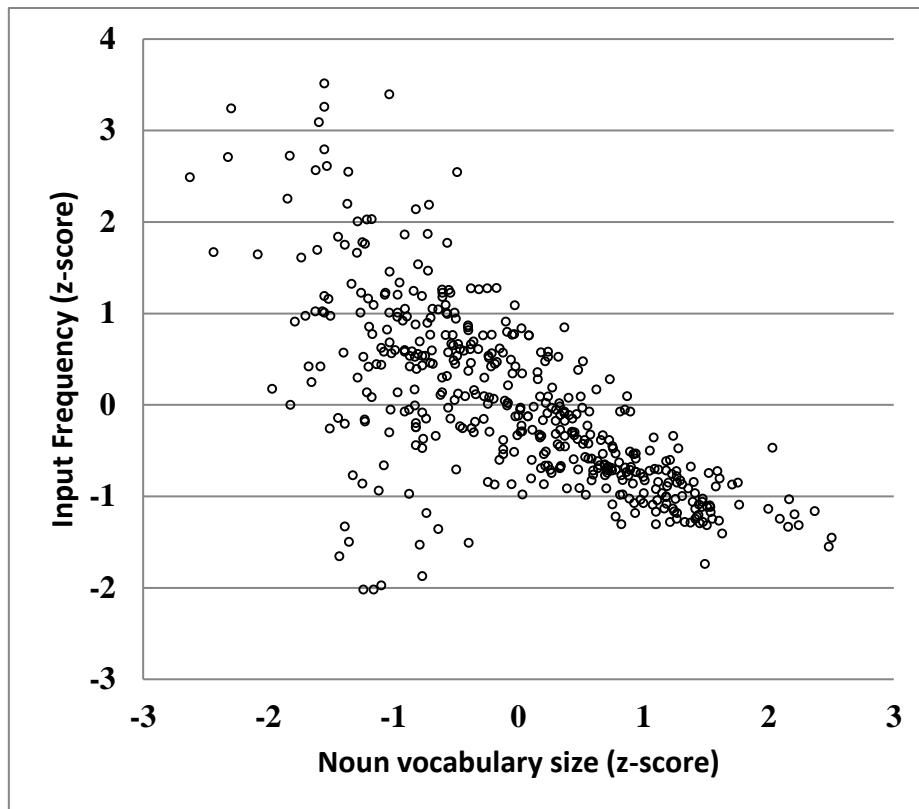


Figure 2: Scatterplot of WF by noun vocabulary size.

As for total vocabulary size, multiple regression analysis was conducted. The variable to predict was in this case noun vocabulary size. The predictors are the same as previously: WF and ND. The model is significant ($F(2, 453) = 398.83, p < 0.01$). WF and ND account for 63.6% of the variance in noun vocabulary size. This time WF and not ND is the strongest predictor according to the t values (table 3). A hierarchical multiple regression showed that WF accounts for 49.4% of the variance in noun vocabulary size ($F(1, 454) = 222.90, p < 0.01$) and ND for 14.2% ($F(1, 453) = 178.00, p < 0.01$).

The same tendency seen above with total vocabulary size is also valid for noun vocabulary size but the effect is stronger. Then, children who present a small noun vocabulary size tend to produce words with high WF and high ND. As noun vocabulary size increases children tend to acquire more words with low frequency and sparse neighbourhood.

Table 3: Table of coefficients for the multiple regression predicting noun vocabulary size

	Standardized coefficients			
	β	t	p	Confidence interval (95%)
(Constant)		0.00	1.00	
ND	-0.44	-13.34	0.00	-0.50 to -0.37
WF	-0.49	-14.93	0.00	-0.55 to -0.42

ND and WF predictors of predicate vocabulary size

As mentioned above, and following the methodology for total vocabulary size and noun vocabulary size, only the 382 children who produced at least 5 predicates were selected for this part.

For predicate vocabulary size, when correlations among variables are investigated, a striking difference appears compared with what was found for noun vocabulary size. No correlation is found between predicate vocabulary size and WF ($r(382) = 0.04$; $p = 0.47$) or predicate vocabulary size and ND ($r(382) = -0.09$; $p = 0.08$). WF and ND are not correlated either ($r(382) = 0.08$; $p = 0.12$). Therefore, WF and ND seem to have no influence on the size of predicate vocabulary unlike what we have seen for the size of noun vocabulary.

Plots of predicate vocabulary size and ND and predicate vocabulary size and WF respectively are shown in Figures 3 and 4. The two plots show that no relationship can be found between the size of predicate vocabulary and ND or the size of predicate vocabulary and WF, although we can mention a high variability for both ND and WF for low vocabulary size.

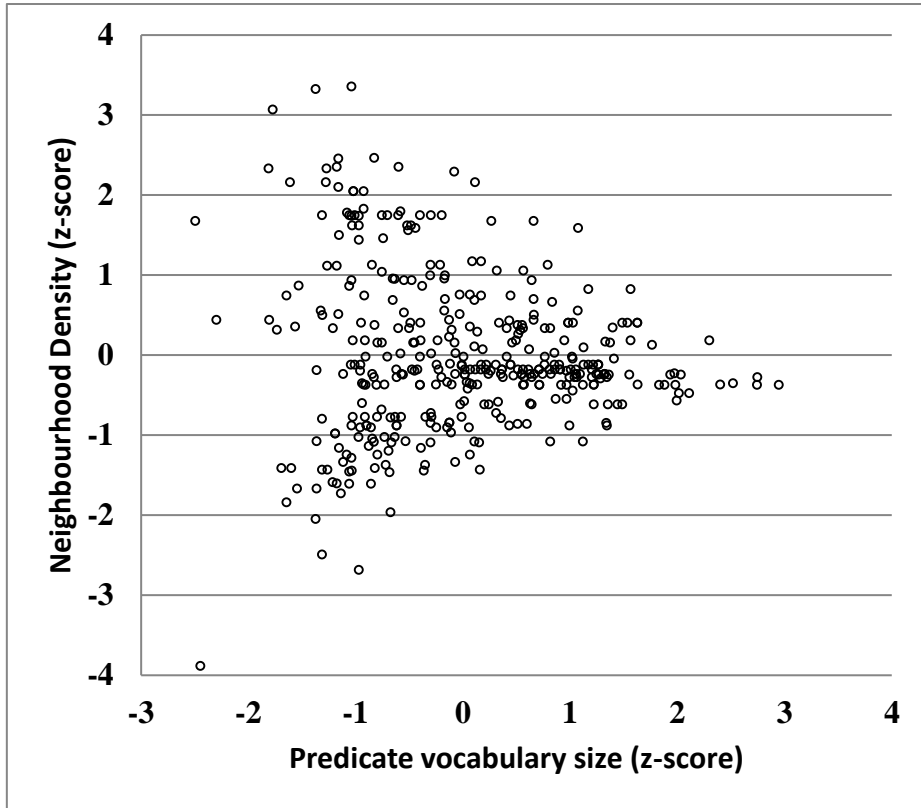


Figure 3: Scatterplot of ND (ND) by predicate vocabulary size.

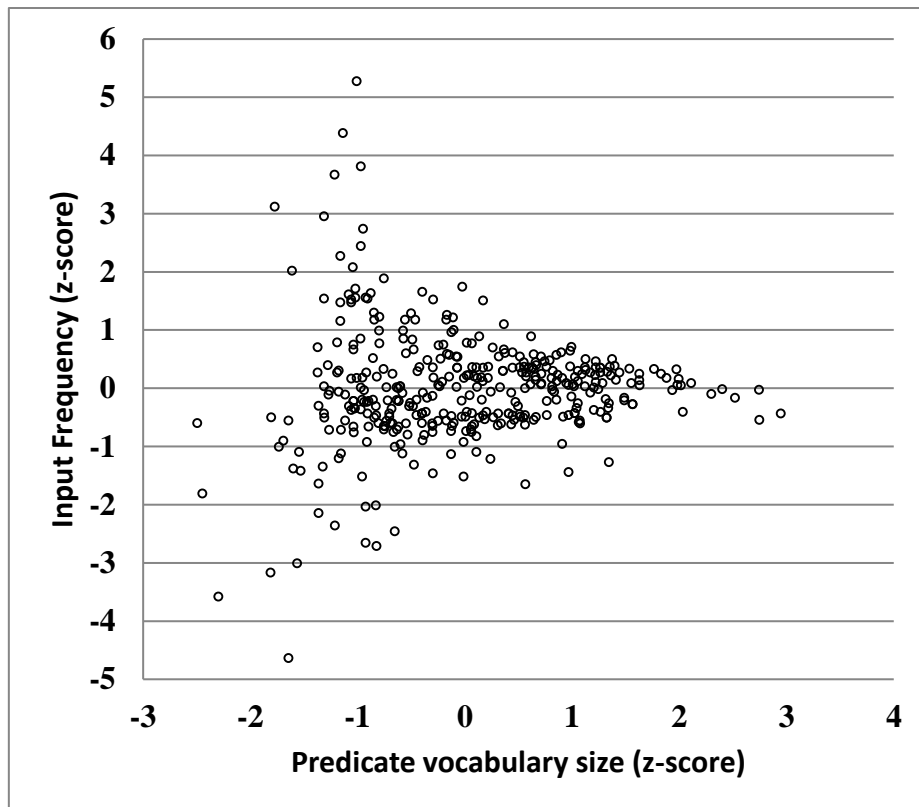


Figure 4: Scatterplot of WF by predicate vocabulary size.

As no correlation was found between predicate vocabulary size and our possible predictors (WF and ND), no multiple regression analysis was attempted with predicate vocabulary size as the variable to predict.

Discussion

We have seen that ND and WF correlate negatively with the size of total vocabulary, predicting 45% of its variance. Children with low vocabulary size tend to have high ND and high WF words in their vocabulary. Subsequently, the question was: do we find the same trend within different grammatical categories? We found a strong negative correlation of ND and WF for the size of noun vocabulary predicting 63.6% of its variance, whereas no influence was found for the size of predicate vocabulary. We can conclude that the influence

of ND and WF found when we looked at total vocabulary size was mainly due to the noun category.

The results provide additional evidence that early acquired words, at least nouns in that study, reside in dense neighborhoods whereas later acquired words reside in sparse neighborhoods. It mirrors results obtained in previous studies (Hollich et al., 2002; Storkel, 2001, 2004). It provides as well additional evidence that early acquired words (nouns in that study) are more frequent ones than later acquired ones.

We found that WF correlates negatively with the size of noun vocabulary. It is the opposite of what was found by Stokes *et al.* (2012a) for French. We replicated the work done by Stokes *et al.* (2012a) using the same data but computing medians instead of mean. The only change between these two studies concerns the choice of using medians instead of means. This choice was necessary due to a large and unequal dispersion of WF of the monosyllabic words ($m = 251.75$; $SD = 1158.84$; $median = 47.31$). We found that WF was, this time, negatively correlated with vocabulary size ($r(208) = -0.67$; $p < 0.01$).

The fact that WF is negatively correlated with the size of noun vocabulary is in line with previous work (*e.g.* Storkel 2009, Goodman *et al.* 2008). It is also in line with Stokes *et al.* (2012b) who studied the influence of WF, ND, and word length in Danish. In this last study, Stokes *et al.* (2012b) used the mean as in previous work. The given hypothesis for the negative correlation between vocabulary size and WF for Danish is the fact that they only included a few verbs (4), compared to English or French analysis, due to the Danish morphology of verbs. As you can see in Table 4, in French, nouns and predicates don't show the same WF distribution and predicates represent 40.5% of FCDI selected words. Another possible hypothesis is the fact that in Danish the dispersion of WF is limited (SD value is approximately one-quarter of the mean) but not in French (SD value is more than four times higher than the mean, see table 4)

Table 4: WF Mean and SD for different CDI dataset

CDI	Danish	French	French nouns	French predicates
Mean WF	99.89	251.75	66.94	523.77
SD WF	24.29	1158.84	88.63	1790.18

Using the median for French allowed us to avoid this dispersion problem and to find results in line with previous studies. However, one of the main results of Stokes *et al.* (2012a, 2012b) is partially confirmed here. ND is a predictor of vocabulary size but only for nouns. The first nouns acquired by children have a dense neighborhood. This factor is not a predictor for the size of predicate vocabulary. This finding may be due to the fact that predicates in French are in general acquired later than nouns. At that time children may have already changed their learning word strategy or other factors linked to predicate can play a role (*e.g.* concreteness, syntactic complexity, informational load...).

The results of our study fit perfectly with the emergentist coalition model (Hirsh-Pasek, Golinkoff & Hollich, 2000) which assumes that a range of cues are available to the child to learn words but that how these cues are used can vary over time. According to our data, the cues can be different according to the grammatical nature of the learned words: ND and WF seem to play a role in the age where nouns are acquired but not in the age where predicates are acquired.

Further research

While this study provides strong insights into the influence of ND and WF on word learning by French infants and toddlers, several limitations are important to keep in mind. The selection of monosyllabic words is according to us its main limitation. However, we made this

methodological choice in order to have comparable data with other languages and studies. It is possible that the results would be different if disyllabic words were included, especially working on French where disyllabic words are more numerous in terms of type than monosyllabic ones (43097 disyllabic words are listed in Lexique 3 and 9509 monosyllabic words) even if, from a frequency perspective, disyllabic words are less frequent than monosyllabic ones (73% of words heard by French monolinguals are monosyllabic and 21% are disyllabic. This fact is mainly due to the high frequency of monosyllabic grammatical words: pronouns, prepositions...). In a pilot study, we tried to find an influence of word phonetic complexity using the IPC scoring (index of phonetic complexity, Jakielski (1998)) on the age of acquisition. Unfortunately, the phonetic complexity of monosyllabic words is highly correlated to ND and almost no additional variation was explained by adding IPC to the model. Including words of more than one syllable should help to disentangle the effect of phonetic complexity and ND.

Finally, three other questions need to be explored. The first question concerns the type of vocabulary. This study only looked at expressive vocabulary. The next step will be to open the analysis to receptive vocabulary and look if we find similar or different results in particular concerning the noun/verb dichotomy. Future work is also needed to further examine how the effect of lexical characteristics such as ND and WF may change over time. Finally The last extension concerns children who continue to have limited vocabulary. ND was mentioned as a factor used by children to learn their first nouns. Do the late talkers, use this strategy longer or do they not learn their first words from dense neighbourhoods like typically developing children do?

Acknowledgments

We take the opportunity to acknowledge Stephanie Stokes (Professor at the University of Canterbury) for having invited us to dive more deeply into the data as well as Anne-Éléonore Lelarge, Master student at the University Lumière-Lyon2 for assistance.

References

- Cameron-Faulkner, T., Lieven, E. & M. Tomasello (2003). A construction based analysis of child directed speech. *Cognitive Science: A multidisciplinary Journal*, 27 (6), 843-873.
- Charles-Luce, J. & P.A. Luce (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17 (1), 205-215.
- Charles-Luce, J. & P.A. Luce (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727-735.
- Coady, J. A., & R. N. Aslin (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30, 441-469.
- Edwards, J., Beckman, M. E. & B. Munson (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47, 421-436.
- Ellis, N. C., 2002. Frequency effects in language acquisition: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143-188.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *MacArthur Communicative Development Inventories: user's guide and technical manual*. San Diego, CA: Singular Press.

Gallaway, C. & B. J. Richards (1994). *Input and Interaction in language acquisition*. Cambridge. 336 p.

Goodman, J.C., Dale, P.S. & P. Li P. (2008) Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35 (3), 515–531.

Gopnik, A. & S. Choi (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language*, 10, 199–215.

Gopnik, A. & S. Choi (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (eds.), *Beyond Names for Things: Young Children's Acquisition of Verbs*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hart, B., & T. Risley (1995). Meaningful differences in the everyday experience of young American children. Baltimore: Brookes.

Hirsh-Pasek, K., Golinkoff, R.M. & Hollich, G. (2000). An emergentist coalition model for word learning: mapping words to objects is a product of the interaction of multiple cues. In *Becoming a word learner*. New York, NY: Oxford University Press.

Hollich, G., Jusczyk, P. W., & Luce, P.A. (2002). Lexical neighbourhood effects in 17-month-old word learning. In B. Skarabela, S. Fish, & A.H.-J. Do (Eds), *Proceedings of the 26th Annual Boston University Conference on Language Development* (Vol1). Sommerville, MA: cascadilla

Jakielski, K. J. (1998). Motor organization in the acquisition of consonant clusters. PhD thesis, University of Texas at Austin. Ann Arbor Michigan, UMI Dissertation services

- Kern, S. & F. Gayraud (2010). « *L'IFDC* », Editions La Cigale, Grenoble.
- Klee, T. & C. Harrison (2001). *CDI Words and Sentences validity and preliminary norms for British English*. Paper presented at Child Language Seminar, University of Hertfordshire, England.
- Landauer T. K. & L. A. Streeter (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 12:119–13.
- Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120, 2546–2556
- Logan, J. S. (1992). A computational analysis of young children's lexicons (Research on Spoken Language Processing Technical Report No. 8). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Maekawa, J. & H. L. Storkel, (2006). Dynamic assessment and word learning. *ACQuiring Knowledge in Speech, Language, and Hearing*, 8, 103-105.
- New, B., Brysbaert, M., Veronis, J., Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661-677
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., Slowiaczek, L.M. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, 4, 75–95.
- Rice, M. L., Oetting, J. B., Marquis, J., Bode J., Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, 37:106–122.
- Snow, E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4, 1-22.

- Stokes, S. (2010). Neighborhood Density and Word Frequency Predict Vocabulary Size in Toddlers. *Journal of Speech, Language, and Hearing Research* Vol.53 670-683 June 2010.
- Stokes, S., Kern, S. & dos Santos, C. (2012a). Extended Statistical Learning as an Account for slow vocabulary growth. *Journal of Child Language*, 39, 105-129
- Stokes, S. F., Bleses, D., Basboll, H., Lambertsen, C. (2012b). Statistical Learning in Emerging Lexicons: The Case of Danish. *Journal of Speech, Language, and Hearing Research*, 55-5, 1265-1273.
- Storkel, H.L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language & Hearing Research*, 44, 1321-1337.
- Storkel, H.L. (2004a). Do children acquire dense neighbourhoods? An investigation of similarity neighbourhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201-221
- Storkel, H.L. (2004b). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighbourhood density. *Journal of Speech, Language, and Hearing Research*, 47, 1454-1468.
- Storkel, H.L. (2009). Developmental differences in the effects of phonological, lexical, and semantic variables on word learning by infants. *Journal of Child Language*, 36, 291-321.
- Storkel, H. L. & M. A. Rogers (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics*, 14, 407-425.
- Vitevitch, M.S. (2002). The influence of phonological similarity neighbourhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 735-747.

Vitevitch, M. S. & M. S. Sommers (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, 31, 491-50

Vitevitch, M. S. & M. K. Stamer (2006). The curious case of competition in Spanish speech production. *Language & Cognitive Processes*, 21, 760-770.

Vihman, M. M. (1996). *Phonological Development: The Origins of Language in the Child*. Cambridge, MA: Blackwell.

Weizman, Z.O. & C.E. Snow (2001). Lexical output as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology*, 37, 265–79.

Appendix

Population: 462 children who produced at least five words or more

Age	# of children	TV	Mean of Mean-WF.	Mean of Med-WF	Mean of Med-Log-WF	Mean of Mean-ND	Mean of Med-ND
16	11	13.00 (9.51)	147.38 (40.44)	92.69 (33.27)	1.93 (0.19)	26.98 (2.35)	27.41 (4.12)
17	10	23.20 (24.84)	172.24 (82.79)	90.74 (29.03)	1.94 (0.13)	29.21 (3.24)	30.10 (4.03)
19	10	49.50 (34.37)	179.79 (62.26)	84.07 (24.17)	1.91 (0.11)	25.86 (1.71)	26.40 (2.90)
20	28	50.96 (35.60)	174.92 (93.64)	73.59 (18.79)	1.85 (0.10)	25.93 (2.73)	25.73 (4.04)
21	49	49.90 (37.22)	193.61 (147.07)	74.70 (24.09)	1.86 (0.12)	25.98 (2.66)	26.11 (3.57)
22	39	61.54 (41.92)	198.49 (102.04)	72.29 (20.42)	1.84 (0.11)	25.74 (2.74)	25.18 (3.77)
23	35	64.71 (43.99)	198.11 (87.25)	77.91 (28.70)	1.87 (0.12)	25.52 (1.93)	25.37 (3.39)
24	88	78.78 (50.40)	190.70 (97.88)	69.64 (27.62)	1.82 (0.12)	25.40 (2.27)	25.20 (3.36)
25	40	105.80 (50.45)	206.03 (83.56)	63.04 (11.96)	1.79 (0.08)	24.35 (1.40)	24.20 (2.32)
26	30	98.87 (57.52)	197.14 (83.55)	68.44 (19.26)	1.82 (0.10)	24.49 (1.31)	24.10 (2.33)
27	24	116.75 (51.84)	232.81 (78.94)	63.85 (13.03)	1.80 (0.08)	24.46 (1.12)	24.00 (1.55)
28	34	126.38 (48.53)	244.08 (88.77)	63.24 (14.39)	1.79 (0.09)	24.02 (1.15)	23.06 (1.55)
29	35	149.77 (45.22)	258.75 (66.60)	59.18 (9.17)	1.77 (0.07)	23.71 (0.87)	23.07 (0.47)
30	29	142.45 (51.48)	229.26 (81.23)	57.06 (8.45)	1.75 (0.07)	23.87 (0.98)	23.24 (1.44)
All	462	86.60 (57.73)	205.04 (97.55)	69.58 (22.46)	1.83 (0.11)	25.14 (2.24)	24.86 (3.21)

Mean (Standard Deviation)

Age in months

TV: Total vocabulary

Med: Median

WF: Word Frequency given per million words