

Accepted for publication
Journal of Speech, Language, & Hearing Research, 6th Sept 2019

Phonology, Semantics, and the Comprehension-Expression Gap in Emerging Lexicons

Stephanie F. Stokes¹, Elise de Bree², Annemarie Kerkhoff³, Mohammad Momenian¹ and Tania

Zamuner⁴

Department of Speech and Hearing Sciences, University of Hong Kong, Hong Kong¹,
College of Child Development and Education, University of Amsterdam, Amsterdam, The

Netherlands²,

Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands³, and

Department of Linguistics, University of Ottawa, Ottawa, Canada⁴

Corresponding author: Stephanie F. Stokes, sstokes@hku.hk, Division of Speech and Hearing Sciences, University of Hong Kong, MW749, Bonham Road, Hong Kong, Telephone: +852 95394047

Conflict of Interest: There are no conflicts of interest

Funding: We did not receive funding for this work

Key words: lexical development, phonological density, semantic density, comprehension-expression gap

Abstract

Purpose: Children come to understand many words by the end of their first year of life, and yet, generally by 12 months only a few words are said. In this study, we investigated which linguistic factors contribute to this comprehension-expression gap the most. Specifically, we asked: are phonological neighborhood density (PND), semantic neighborhood density (SND), and word frequency (WF) significant predictors of the probability that words known (understood) by children would appear in their spoken lexicons?

Method: Monosyllabic words in the active (understood and said) and passive (understood, not said) lexicons of 201 toddlers were extracted from the Dutch Communicative Development Inventory (NCDI) parent-completed forms. A Generalised Linear Mixed Effects Model (GLMEM) was applied to the data.

Results: PND and WF were independently and significantly associated with whether or not a known word would be in children's spoken lexicons but SND was not. There were individual differences in the impact of WF on the probability that known words would be said.

Conclusion: The novel findings reported here have two major implications. First, they indicate that the comprehension-expression gap exists partly because the phonological distributional properties of words determine how readily words can be phonologically encoded for word production. Second, there are likely subtle and complex individual differences in how and when the statistical properties of the ambient language impact on children's emerging lexicons that might best be explored via longitudinal sampling of word knowledge and use.

By 12 months of age the average toddler is reported to understand about 78 words, and can gesture towards and locate named objects in his/her environment; however, at the same time the average child is reported to say only six words (Frank, Braginsky, Yurovsky, & Marchman, 2016; Schneider, Yurovsky, & Frank, 2015). By 18 months the counts increase to 262 words understood and 57 said, and the gap remains. Words occur in a child's passive lexicon, meaning that they are understood but not said, for some months before they occur in the active lexicon, meaning they are said as well—this has been called an early comprehension-expression gap (Schneider, et al., 2015). Why does the comprehension-expression gap exist? Previous work has discussed the role of age-dependent factors such as articulatory control or limited working memory as a mechanism for explaining the lag in production (Majorano, Vihman, & DePaolis, 2014). However, the influence of articulatory control has traditionally been viewed as minor, as children can often imitate avoided structures (Menn & Matthei, 1992). That is, because children can imitate phonological structures that they do not say in spontaneous speech, it does not seem to be the case that articulation control per se explains the lag in production. Word learning—mapping sound to meaning—is a dynamic process that involves encoding and consolidating words in long-term memory (Munro, Baker, McGregor, Docking, & Arculi, 2012). Early on, learning is characterized by sudden starts and stops, with unstable lexical representations (Bloom, 1974). Word learning studies show that comprehension and production of novel words (e.g. *pib*, *booma*) may fluctuate from week to week (Gershkoff-Stowe & Hahn, 2013). Still, the advantage for comprehension is always found in such 'fast mapping' studies. For instance, in Horst and Samuelson's (2008) study, children recognized word-object pairings after two exposures but were unable to name the objects. Words may first be produced only when cued or prompted, or when the word has been

Running head: COMPREHENSION-EXPRESSION GAP

uttered by an adult relatively recently, only gradually progressing to the more demanding stage of free recall (Munro et al., 2012). Generally, while there are multiple verbal and nonverbal cues to guide comprehension, retrieval for production is considered more taxing (Gershkoff-Stowe & Hahn, 2013). The comprehension-expression gap may thus reflect the different processing demands for lexical access (sound-to-meaning mapping) and production (meaning-to-sound mapping), with the latter arguably requiring more detailed phonological representations (Huttenlocher, 1974, cited in Gershkoff-Stowe & Hahn, 2013). Lexical access accounts can also explain why a similar comprehension-expression gap is found in adults, both in second language acquisition and novel word learning studies (e.g. Gershkoff-Stowe & Hahn, 2013). However, as yet little is known about the exact nature of the factors underlying the (early) comprehension-expression gap.

One pathway to answering the question is to ask whether there are qualitative differences between the words in children's receptive and expressive lexicons. If so, in what ways? To answer the question a repository of children's lexicons would be required, and those lexicons would need to be separated into words that are in the child's receptive lexicon but are not said (passive lexicon), and words that are in the receptive lexicon and are also said, an expressive lexicon (active lexicon). To illustrate our point, a concrete example may be useful. In the data we report in the current study, on a parent-report checklist of child vocabulary, one child had a total receptive lexicon of 305 words, and said 139 of them, i.e., there were 139 words in his active (expressive) lexicon. This means the child understood 166 words that he did not say - these words would be counted in his passive lexicon. We see then that the terms 'active' and 'expressive' lexicons have the same meaning, and will contain the same number of words, but the 'passive' and 'receptive' lexicons are not the same.

To date, we know of only two studies that asked whether there are qualitative

Running head: COMPREHENSION-EXPRESSION GAP

differences in the statistical properties of the words that children say and understand (Sahni & Rogers, 2008; Stokes, 2014). The first (Sahni & Rogers, 2008) reported on the statistical properties of words in the receptive and expressive lexicons of children aged 8-30 months. They used lexical norms from the Lex2005 database inventories (Dale & Fenson, 1996) and three sets of words from the MBCDI checklists (MacArthur-Bates Communicative Development Inventory; Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, 1993), i.e. Infant Knows, Infant Says (both 8-16 months) and Toddler Says (16-30 months). Then they assessed the contribution of different variables for both the receptive and expressive lexicon. The first variable was phonological neighborhood density (PND), a measure of the phonological characteristics of words. Specifically, PND refers to the number of words that differ from a target word by a single phoneme (Charles-Luce & Luce, 1990). It should be noted, however, that Sahni and Rogers (2008) adopted a different metric of PND, namely Joanisse and Seidenberg's (1999) PND metric of degree of shared binary distinctive features (e.g., voiced/voiceless). Sahni and Rogers also included the input frequency of the words. Finally, they included semantic neighborhood density (SND), the conceptual similarity between words. They reported that PND and input frequency predicted the age of acquisition of words in the expressive lexicon, and that SND and input frequency predicted the age of acquisition of words in the receptive lexicon.

The PND effect for expressive lexicons was first reported by Storkel and colleagues (e.g., Storkel, 2001; Storkel, 2004a) and has since been replicated (Carlson, Sonderegger & Bane, 2014, Storkel, 2009; Stokes, 2010; Stokes, Bleses, Basbøll & Lambertsen, 2012a; Stokes, Kern & dos Santos, 2012b; Stokes, 2014), but Sahni and Rogers (2008) was the first report of a SND effect on children's receptive lexicons. This result raises the intriguing propo-

sition that semantics determines which words a child comes to understand, but phonology determines which words are first uttered.

To investigate this proposition, the child's receptive lexicon must be parsed into those words that are only understood (passive) and those that are understood and said (active). This strategy was used by Stokes (2014), who conducted the second study known to us on differences between the statistical properties of the words that children understand and say compared with the words children understand but do not say. Stokes proposed that the phonological factor of PND was implicated in the comprehension-expression gap. She claimed that at the outset of word learning, children's phonological representations are strong enough to allow for the recognition and comprehension of a word, but are not strong enough to support the production of the same word. Evidence for this occurs in fast mapping tasks where children are asked to identify a newly-named novel object from an array of novel objects. Toddlers can find the named novel object from a choice of objects (Stokes & Klee, 2009), and preschool children (3-6 years) can choose the correct novel word form from a choice of novel word forms in an identification task (e.g. *dorb*, *vorb*, *zinnip*; Gordon, McGregor, Waldier, Curran, Gomez & Samuelson, 2016) but cannot produce it when asked to recall the label (Munro, Baker, McGregor, Docking, & Arciuli, 2012; Stokes & Klee, 2009). That is, children easily recognized the novel word forms, but could not generate them. Stokes (2014) argued that the PND of a word was a strong determining factor as to whether or not those identified novel words were likely to be said. She claimed that words from dense phonological neighborhoods in the ambient input provided multiple opportunities for children to hear recurring phonological strings (consonants and vowels shared across neighbors), far in excess of heard forms for sparse words, allowing children to build more robust phonological representations for dense than sparse words.

The Phonological Density Advantage in Emerging Lexicons

Words from dense neighborhoods are those with many phonological neighbors (i.e., *cat*, with neighbors such as *hat, mat, fat, cut, cap*), whereas sparse words have few (i.e., *foot*). In computational models of language production (Gordon & Dell, 2001), dense words benefit from increased activation as a function of spreading activation when a member of the neighborhood is activated for word production. Gordon and Dell described a bidirectional interaction effect whereby phonological density facilitates both lemma and phoneme retrieval. Retrieval of a lemma for production stimulates the activation of phonological neighbors through shared phonemes, which in turn provide feedback that strengthens the phonological representations of both the target lemma and target phonemes. Sparse words have few neighbors and do not have the same extent of facilitated activation. This spreading activation means that saying phonologically dense words is easier (faster naming and fewer errors) than saying phonologically sparse words, as has been reported (Vitevitch, 2002).

This supposition has recently been supported by a computational study (Takáč, Knott, & Stokes, 2017). Takáč et al. (2017) trained a simple recurrent network (SRN) on the phonological form and meaning of 268 monosyllabic English MBCDI (Fenson, et al., 1993) words understood by toddlers. The model was also exposed to the phonological forms of 2320 monosyllabic words without meaning, representing unknown words. Input frequency to the SRN was controlled to match the normal frequency of occurrence of these words in the ambient language (English). Words were coded for PND, biphone frequency within words, word frequency, and word length. The SRN was tested on a word production task in which a word form was generated for each meaning. The learning pattern resembled the variability seen in children, as the first-learned words were dense words (that is, the SRN was faster at learning

Running head: COMPREHENSION-EXPRESSION GAP
words with high PND). Presumably this occurs in children because of a facilitative neural spreading activation effect for dense words (Gordon & Dell, 2001).

This computational study reflected the effect of PND seen in actual children's emerging lexicons. The active lexicons of children younger than 30 months contain words of a higher mean PND than their passive lexicons (Stokes, 2014). While children in Stokes's study understood both phonologically sparser (lower PND) and denser (higher PND) words, denser words were more likely to be said, arguably because these words share recurring phoneme strings with many other words in the ambient language, and sparser words do not. Recurrence of these phoneme strings both in the ambient input and the child's own productions is believed either to reduce the working memory demands for word production or increase the speed of activation of denser words, or both. In other words, at the same point in development, typically developing children's passive and active lexicons contain both dense *and* sparse words but there are more dense words than sparse words in their active lexicons.

Beyond the Density Advantage

Stokes (2014) claimed that it is the phonological density of words that determines whether or not an understood word is said by the child, at least at the stage when a word emerges in speech production. However, this is the only study so far to have examined the difference between active and passive lexicons. (Recall that Sahni & Rogers (2008) examined the receptive and expressive lexicons, not passive and active lexicons.) Furthermore, Stokes's study only examined PND as a contributing factor, whereas other linguistic factors are likely to influence the passive and active lexicons as well, suggesting the need for further research. If phonological factors are the strongest in accounting for the comprehension-expression gap then Stokes's theory of a phonological encoding gap between knowing and saying a word would be supported. In the present study, we assess to what extent two other linguistic factors contribute

to the comprehension-expression gap. Factors under examination in addition to PND are word frequency and semantic neighborhood density. We did not include the factor of phonotactic probability in our analyses because it has not been identified as a significant predictor of emerging lexicons beyond age, word length, word frequency, and neighborhood density (Carlson et al., 2014; Maekawa & Storkel, 2006; Zamuner & Thiessen, 2018). Similarly, we did not include word length as a factor because although we know that words that occur first in spoken lexicons are short, having fewer phonemes than later-produced words (e.g., Storkel, 2009, Stokes et al., 2012a), word length has been shown to be less important than PND as a predictor of lexical development in toddlers. Storkel (2004b) reported that PND, but not word length in phonemes, was a significant predictor of the age of acquisition of words in children aged 8-30 months, and Stokes et al. (2012a) found that PND accounted for 39% of the variance in vocabulary size in Danish-speaking toddlers, with word length contributing 2.2% of additional unique variance accounted for. As word length has not been a significant predictor beyond PND, it is not considered further here.

Word frequency. The frequency of occurrence of a word in the ambient language must be related to the frequency with which the child is exposed to the semantic and phonological properties of that word. A word of low PND, a sparse word, may be highly frequent in the ambient language so that regardless of low PND that word may appear in the active lexicon early in lexical development. This question has been addressed previously with respect to MBCDI spoken lexicons, but not passive lexicons (Stokes et al., 2012b, Stokes, 2014; Carlson et al., 2014; Jimenez & Hills, 2017). In prior studies more frequent words were learned earlier.

Stokes (2014) has argued that the comprehension-expressive gap is most apparent for phonologically sparse words because they have few phonological connections with other words and therefore do not benefit from the spreading neural activation that is apparent for phonolog-

ically dense words during activation for production. The claim is that the words in the passive lexicon do not pass easily into the active lexicon because they lack the benefit of spreading neural activation during production attempts. If the lexical characteristic of word frequency has an impact on the active lexicon then we would expect that an increase in word frequency would increase the probability that words would appear in the spoken lexicon. Based on prior research, we claim that dense and more frequent words present many opportunities for the child to retrieve, formulate and generate frequent onsets and rhymes that are shared among neighbors and therefore denser words become active sooner than do sparser words in emerging lexicons. Thus, children will be more likely to say dense words than sparse words. That is a higher PND should increase the likelihood that a known word will be in the active lexicon. However, the impact of word frequency has been well established and it is likely that this variable will also play a role in predicting whether known words are in the active lexicon.

Semantic density. In addition to phonological and word frequency cues, semantic factors have also been considered in the search for drivers of early word learning. Semantic density (SD) refers to the conceptual similarity between words (Li, Bandar & McLean, 2003). Calculation of whether or not words are conceptually similar have been variably based on the perceptual similarity of referents (e.g. Howell, Jankowicz, & Becker, 2005, for example size, color, texture), corpora collocation (Evert, 2008, for example *years ago*), semantic set size (association norms, e.g., Nelson, McEvoy & Schreiber, 2004, for example when adults hear the word *cat* they are most likely to say *dog*), and superordinate category (e.g., *toy, food*). Recent research on the impact of semantic density in early word production has yielded equivocal results.

Beckage, Aguilar and Colunga (2015) using Howell et al.'s (2005) perceptual features matrix reported that PND, rather than SND, played an important role in predicting the age of

emergence of children's spoken words in children aged 16-30 months. Sahni and Rogers (2008) for their study of words known by children aged 8-30 months also used the Howell et al. (2005) framework to calculate SND, and reported semantic effects in receptive lexicons, so it is plausible that SND derived from a perceptual feature matrix facilitates the establishment of a passive lexicon. However, in a network analysis of MBCDI noun production norms across time (16-30 months), Hills, Mauene, Maouene, Sheya, and Smith (2009) reported that both phonological and semantic connectedness of words in the learning environment (based on the number of phonological neighbors and noun associations respectively) predicted spoken noun emergence, in conjunction with word frequency. Importantly, it was adult-derived semantic associations, not perceptual features that predicted growth in word production.

There has been no comparison of perceptual feature matrices and word association effects for receptive lexicons meaning that there is no conclusion to be drawn on which measure is the most relevant for studying emerging lexicons. Drawing on Hills et al. (2009), we use association norms (De Deyne & Storms, 2008; De Deyne, Navarro & Storms, 2013) as our SND metric in the current study. To our knowledge, the word association norms measure is the most detailed measure available of semantic associations in Dutch. These word associations are taken to be the most direct way for accessing semantic knowledge (De Deyne et al., 2013; Nelson et al., 2004), as this is a measure that allows for more flexibility in expressing relationships compared with other measures because of multiple associative responses. Including this measure as our SND metric in the current study arguably provides the best chance of identifying semantic influences on emerging lexicons. In summary, Beckage et al (2015), and Sahni and Rogers (2008) reported the impact of PND in emerging spoken lexicons, and Hills et al (2009) reported both phonological and semantic effects on emerging spoken lexicons. Therefore we

predict that PND and word frequency, but not SND will be significantly associated with the probability that known words will be in the active lexicon.

Establishing the Lexicon from the Communicative Developmental Inventory

Establishing passive and active lexicons: Parent report

Prior studies of the PND of children's emerging lexicons have used data collected via linguistic adaptations of the MBCDI (Fenson et al., 1993). Few of the adaptations of the MBCDI Words and Sentences for toddlers, in English or other languages, afford the opportunity to examine the composition of the active and passive lexicons separately. In most toddler versions parents check off words they have heard their child say, which implies that the child also understands the word. This means that there is no option for the parent to indicate words that their child understands, but does not produce. The infant version, Words and Gestures, does make this distinction. Stokes (2014) studied the impact of PND in the active and passive lexicons of children at two ages (18 and 24 months) using a British (Hamilton, Plunkett, & Schafer, 2000) parent checklist of child word knowledge (active and passive), and found that the PND effect was particularly strong at 18 months. Studying input and linguistics factors in the active and passive lexicons of children up to 30 months affords the opportunity to examine if the density advantage diminishes with time or with increasing lexical development. One toddler version that allows for separation of the active and passive lexicons is the Dutch adaptation—The Nederlandse Communicative Development Inventory (NCDI; Zink & Lejaegere, 2002). This adaptation consists of 702 words that parents code as *understands* or *understands and says*, or leaves blank.

While the structure of the NCDI questionnaire provides a way to code for children's active and passive lexicons, it also allows us to examine whether patterns of PND reported for English are also found in a previously unexamined language, i.e. Dutch. Our prediction was

Running head: COMPREHENSION-EXPRESSION GAP
 that Dutch would show similar patterns as English, given that both English and Dutch are Germanic languages, with similar syllable structures. We also predicted similar findings in Dutch based on Ziegler and Goswami (2005), who in a cross-linguistic comparison of the compositions of phonological neighborhoods across English, Dutch, German and French, reported a higher proportion of rhyme neighbors (hat/cat) than consonant neighbors (hat/hit) and lead neighbors (hat/ham), in all languages. Child Dutch lexicons were also found to be composed of a greater proportion of rhyme neighbors (hat, cat) than consonant or lead neighbors (De Bree, Zamuner, & Wijnen, 2004). (Also see Marinus, Nation, and de Jong (2015) and Norris and Kinoshita (2012) for a comparison of PND in English and Dutch.)

The accuracy of parent report of child knowledge.

Using parent checklists of child vocabulary knowledge is not uncontentious due to questions about the accuracy of parent report (e.g., Houston-Price, Mather & Sakkalou, 2007). Fenson, Dale, Reznick, Bates, Thal and Pethic (1994) reported that parents are reliable in their test-retest estimates of the number of words their child understands (MBCDI infant form) or says (MBCDI infant and toddler forms). Pearson correlations on a test-retest trial on the infant form for word comprehension was $r(135) = .87$, and for word production it was $r(135) = .95$. They also reported a test-retest Pearson correlation of $r(214) = .96$ for word production for the toddler form.

The test-retest reliability may be weaker in terms of the actual words checked off by parents of children with a language impairment. Yoder, Warren & Biggar (1997) studied the stability of parent report of the CDI Infant Scale across a two-week period in a sample of 17 parents of children with developmental delays. Although there was excellent reliability for the overall number of words understood ($r(15) = .93$), item-by-item stability was modest, with Cohen's kappa ranging from .62 for verbs to .65 for nouns. However, Styles and Plunkett (2009)

reported that 35 parents were accurate on which words the child understands on the MBCDI when compared with child behavior on an intermodal preferential looking paradigm to assess word comprehension. Further, they suggested that parents of 18 month old toddlers “mark as ‘understood’ those items which their infants will be able to identify correctly, with only one presentation, in an unfamiliar environment, in the presence of potentially confusing distracters.....it appears that... parents set a high bar for their infants’ ‘understanding of object names’” (p. 907).

Present Study

This investigation sought to provide evidence for the novel theory that the phonological density advantage (Stokes, 2014) is the dominant factor accounting for the comprehension-expression gap in emerging lexicons by including multiple variables in one investigation. This innovative approach furthers the search for a plausible explanation for why the comprehension-expression gap exists. To explore which word characteristics influence the probability that a known word appears in the active lexicon we asked are PND, SND and word frequency significant predictors of the probability that known words will be in the active lexicon? Word frequency is well established as a predictor of the age of acquisition of words in spoken lexicons, and PND is well established as a predictor of vocabulary size. Here we expected higher WF and PND to be associated with a higher probability that words would be in the active lexicon.

Method

Participants

The parents of 201 Monolingual Dutch-speaking children¹ completed the Dutch adaptation of the MacArthur-Bates Communicative Development Inventory (Nederlandse Communicative Development Inventory, NCDI, Words and Sentences, Zink & Lejaegere, 2002). Par-

¹ Children and subjects are used interchangeably throughout the text

Running head: COMPREHENSION-EXPRESSION GAP

ents were recruited through the local municipality (city, country). The sample thus consists of an urbanized population. All families, except twelve, were residents of this larger municipality. The twelve others also came from the Randstad area, an urbanized part of the Netherlands. We did not record parental education level, as this was not the goal of our research, but the parents involved in the Babylab studies generally have a high level of maternal education (e.g. Kerkhoff, de Bree, de Klerk & Wijnen, 2013). The children had no known language, hearing or behavioral disorders according to parent report. The average child age was 26.79 months (SD = 2.81), with a range of 23 to 32 months. There were 92 girls and 109 boys.

The authors declare that at present and at the time of the study, the XXX (university) endorses the *WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects*, as well as *The Netherlands Code of Conduct for Scientific Practice* issued in 2004 (revised in 2012) by the Association of Universities in the Netherlands (VSNU). No IRB approval was obtained as the XXX institution did not have an ethical assessment committee at the time when the current study was conducted. Ethical approval was granted by the ethics coordinator in the department of the second and third author (XXX university). Parents provided written consent to participate in the study, completed a brief questionnaire about their child's development, and received feedback on the vocabulary outcomes of their child, if desired (by ticking the box on the form). The results were sent through postal mail, as was a colorful child-friendly pencil. If the scores of the children showed reason for concern, the NCDI was re-sent three months later to check for prolonged lower scores. In case of prolonged lower scores, parents were advised to consult their GP or a speech therapist.

Lexicons

The NCDI consists of 702 words that parents code separately as *understands* or

understands and says. Following Storkel's early ground breaking research on the effects of PND on early lexicons, (Storkel, 2001), Stokes and colleagues (2010, 2012a, 2012b, 2014) established the impact of PND on emerging monosyllabic lexicons in English, French, and Danish. Further, generating cross-linguistic evidence for or against a phonological encoding gap between active and passive lexicons requires that studies use similar methods. The current study extends the range of investigation to include semantic neighborhood density and Dutch and therefore replicates the methods employed in prior studies. We restricted our analysis to the 289 monosyllabic nouns, verbs and adjectives on the NCDI. We acknowledge that imposing this experimental control on word length limits the external validity of the results (Storkel, 2004a), because monosyllabic words cannot reflect a child's entire lexicon. For each child the passive lexicon was determined by subtracting the words in the child's active lexicon (says) from the child's *understands* lexicons, leaving only those words that the child understands but does not say (the child's passive lexicon), yielding two mutually exclusive lists of words in the active or passive lexicon for each child. This was done in case parents checked off the same word on both lists, as 'understood' and 'understood and says'. If the word was in the child's 'says' inventory, it would not be in the Passive lexicon.

(Sub)lexical Characteristics of Words

Values for word per million (WF), phonological neighborhood density (PND) and semantic neighborhood density (SND) were assigned to each of the 289 monosyllabic nouns, verbs and adjectives on the NCDI. The WF and PND values were extracted from CELEX (Baayen, Piepenbrock, & van Rijn, 1993), and the SND was derived from a Dutch word association metric (de Deyne & Storms, 2008). The analyses used the values for the (sub)lexical variables from adult lexicons, as done in previous studies (e.g., Hills et al., 2009) but note that de Bree et al. (2014) reported high correlations between the CELEX (adult) and van der Weijer

Running head: COMPREHENSION-EXPRESSION GAP
(infant-input) corpus (van de Weijer, 1998) for the (sub)lexical characteristics of Dutch words.

The PND, SND and WF values were generated for each word in each of the 289 words.

Data Analysis Plan

We used Generalized Linear Mixed Effects Modeling (GLMEM) using lme4 package (<https://cran.r-project.org/web/packages/lme4/>) with R software (R Development CoreTeam, 2008) to analyze the data. (G)LMEM is a powerful tool which allows researchers to take into account individual differences by modeling random effects in the data enabling a more comprehensive and reliable presentation of the findings (Gordon, 2019; Harel & McAllister, 2019).

The practice suggested by Barr, Levy, Scheepers, and Tily (2013) and Bates, Kliegl, Vasishth, and Baayen (2015) was followed to fit the models. We will explain this practice later in the text. Our dependent variable was binomial (1 vs. 0, where 1 = in the active lexicon, and 0 = in the passive lexicon), showing whether a known word belonged to the active or the passive lexicon (parents also indicated words that their child neither understood nor said and these were marked as unknown and not included in the analysis). The fixed effects were Word Frequency (WF), Phonological Neighborhood Density (PND) and Semantic Neighborhood Density (SND).

Before modeling, we standardized the independent variables using the Z score formula to address scale differences among the variables and avoid model convergence issues. We then checked the variance inflation factor (VIF) for the independent variables in the model. A VIF larger than 5 or 10 is a sign of multi-collinearity in the data (Craney & Surles, 2002). All VIFs were below 2 which showed that multi-collinearity issues did not exist in the data.

We started with a maximal model (Barr et al., 2013; Gordon, 2019) which was informed by the design of our study. In this maximal model, we included all fixed variables (WF, PND, SND) and their interaction terms plus random intercepts of items and subjects together

Running head: COMPREHENSION-EXPRESSION GAP

with by-subject random slopes for WF, PND and SND. Having fit the maximal model, we sought the most parsimonious random effects structure for the data (see Bates et al., 2015 for a discussion on parsimonious mixed models). In order to discover which random effect was not contributing substantially to the model, we did a Principal Component Analysis (PCA; Bates et al., 2015). Informed by the PCA, we dropped the random effects with the lowest variance one at a time and compared the output with the previous full model using the Likelihood Ratio Test (LRT). If the LRT was significant, we selected the model with the lowest AIC (Akaike Information Criterion). We did this procedure sequentially until we found the most parsimonious random effects structure which best explained the covariance structure of our data.

In the second stage, we determined which of the fixed variables had a significant effect using Satterthwaite approximations (Kuznetsova, Brockhoff, & Christensen, 2017). There are many approaches available to determine which fixed effects could be significant (Baayen, Davidson, & Bates, 2008; Barr et al., 2013; Luke, 2017, Pinheiro & Bates, 2000). The Wald test and the LRT are mostly used since they are less computationally expensive (Luke, 2017). However, the application of LRTs on fixed effects has proved anti-conservative (Pinheiro & Bates, 2000). We adopted Satterthwaite approximations to calculate denominator degrees of freedom. This method results in more acceptable Type 1 error rates when compared with Wald and LRT (Luke, 2017). The R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) was used.

Results

The children's mean active lexicon size was 184.92 words ($SD = 78.50$), with a range of 1-289, and the mean passive lexicon size was 48 words (46.49), with a range of 0 - 268.

Do PND, SND and WF significantly predict the probability that a known word will be in the active lexicon?

Running head: COMPREHENSION-EXPRESSION GAP

We followed the exact procedure outlined above. Based on the PCA results and the variance-covariance matrix, we removed those random effects with the lowest variance followed by LRTs. The removal of by-subject SND ($\chi^2(3) = 2.14, p = 0.54$) and by-subject PND ($\chi^2(4) = 7.95, p = 0.93$) random effects did not have any significant effects on the model. The exclusion of by-subject WF ($\chi^2(2) = 60.39, p < 0.001$), however, changed the model fit significantly (see Figure 1 for the variation in WF, PND, and SND slopes among children). The final random effects structure, therefore, included the random intercepts of item and subject, plus the by-subject random slope for WF. By modelling random effects in our data, we took into account the dependency in our data (see Figures 1 and 2). This means the effects we present for our fixed or predictor variables are informed by the covariance structure in our random effects.

Figures 1 and 2 about here

The significant by-subject slope for WF means that there are individual differences in the effect of word frequency. Figure 1 shows the effects of WF, PND, and SND for every individual child in the study. This indicates that the effect of WF was more variable ($SD = 0.21$) than that of PND or SND across children. Although this effect is not comparable to what we see in the intercepts for subjects and items (Intraclass Correlation Coefficients (ICC) for subjects = 0.79, ICC for items = 0.55), it should still be taken into account to give a more reliable interpretation of WF effects. Marginal R^2 (of Fixed effects) and conditional R^2 (of Fixed and Random effects combined) were 0.012 and 0.653 which is a clear indication of how modeling random effects could explain more sources of variance in our data (Johnson, 2014; Nakagawa, Johnson & Schielzeth, 2017).

Within this background of random effects, we examined which variables had independent effects on the probability of a word being in the active lexicon. Of all three variables and

their interactions, only WF and PND had substantial effects (See Table 1 for a summary of the mixed model). That is, the higher the WF and PND values, the higher the likelihood of a word being in the active lexicon (see Figure 3). The random slope of WF significantly improved model fit. This means that the main effect of WF we see in Figure 3 would be slightly different from one child to another ($SD = 0.21$). For example, if we randomly pick two children from the study, we may not see the same exact slope for WF in these two children (Figure 1 shows the WF effect across individual children). However, for PND and SND the slopes in Figure 1 would be more similar among children suggesting these variables may be less subject to change because of individual differences.

Table 1 and Figure 3 about here

Discussion

Why does the early comprehension-expressive gap exist? Why do toddlers understand (know) many more words than they say? Are there distributional properties of words that determine whether or not a known word is said? By separating the active and passive lexicons we were able to explore whether the phonological, semantic and frequency properties of known words may partly determine whether or not those words are said. We hypothesized that PND and WF would be significant predictors of known words being said, and SND would not. Our hypothesis was confirmed because while WF and PND had significant effects in the model, SND did not. The higher the PND and WF value of words, the higher the probability that words would be spoken. This confirms that while the child knows words from denser and sparser networks of the ambient language and less and more frequent words, s/he is more likely to say denser and more frequent words.

The finding of a robust PND effect supports our theory that words in children's active lexicons words come from phonologically denser portions of the ambient language—the learn-

Running head: COMPREHENSION-EXPRESSION GAP
ing environment (Coady & Aslin, 2003) - a learning mechanism named ‘preferential attachment’ by Hills et al. (2009). Children at the earliest stages of establishing an active lexicon say denser words rather than sparser words, regardless of whether or not these children know sparser as well as denser words.

The question is why children say denser words even if they understand both sparser and denser words. There is evidence that denser words benefit from increased neural activation as a function of spreading activation (Gordon & Dell, 2001; Vitevitch, 2002). This would mean that the phonological encoding (assembly) of denser words for production is faster and easier than the assembly of word forms for sparser words. This has been demonstrated many times in naming studies with adults and in some work with older children. Takáč et al. (2017) demonstrated that a computational model is better able to predict the strings of phonemes that co-occur in denser than sparser words. Word assembly for spoken production can be construed as a predictive task where denser words are easier to assemble than sparser words because more frequently encountered word-level phoneme strings become more predictable. This occurs implicitly in word learning as a function of the statistical properties of words in the ambient language. It seems then that the comprehension-expressive gap may in part be due to the child assembling some word forms more readily than others for motor production, as a function of implicit learning.

The WF effect on word learning has been well documented, as has individual differences in the effect of WF on word processing in adults (Brysbaert, Mandera, & Keuleers, 2018). However, the individual differences in the WF effect on whether or not children say known words is a novel finding. In the types of data sets we are analyzing, there may be a more complex interaction between WF and PND. Or it may be the case that for individual children there is a complex relationship between PND and WF. For example, for some children dense

words may be said over sparse words regardless of word frequency, or only in frequent words (Storkel, 2004a). Or there could be a complex relationship between WF, PND, lexicon size and learning over time. Further research is required that systematically tests these relationships in longitudinal studies.

As expected, SND was not a significant predictor of the probability that known words would appear in the active lexicon. Simply put, at this stage of vocabulary acquisition, SND was not an important factor in learning to say words. Sahni and Rogers (2008) reported an SND effect in receptive lexicons but the receptive lexicon in their study included words that were understood and said as well as words that were understood and not said. Parsing the receptive lexicon into active and passive lexicons has allowed us to demonstrate for the first time that while PND and WF had an impact on whether or not known words are in the active lexicon, SND did not. There are two plausible explanations. The first is that demonstrating an impact of SND may be dictated by the way in which SND is conceptualised and the values for words are calculated. We return to this issue in the section on study limitations. The second is that the impact of PND, SND and WF may shift as a function of increasing language development (Storkel, 2009).

Implications for lexical acquisition by Typical and Atypical developers

Our theory is that the early comprehension-expressive gap is not only due to the memory processing demands for lexical access, but may also reflect phonological factors: although a child can come to understand words regardless of their phonological structure, saying a word requires a stronger phonological representation for that word. We have witnessed the impact of phonological neighborhood density most profoundly in the lexicons of children who are slower at learning to talk than their same-aged peers, often referred to as late talkers (Stokes and colleagues, 2010, 2012a, 2012b, 2014). Frank et al. (2016) reported that by 12 months 50% of

Running head: COMPREHENSION-EXPRESSION GAP

American-English speaking infants understand 78 words and say 6 words. If a toddler's passive lexicon contains at least 78 words then one might expect that s/he is starting to talk.

MBCDI norms from the WordBank website (Frank et al., 2016) are useful for comparing a child's lexicon size with her/his peers' lexicons. Children whose spoken lexicon size is at or below the 10th percentile for their age on the MBCDI can be identified as late talkers. Late talkers tend to say words of high phonological neighborhood density, regardless of the language studied. Forty percent of these children who have few words at two years continue to have a language delay by four years (Dale, Price, Bishop & Plomin, 2003). Increasing children's spoken lexicons may be achieved by early language intervention that focuses on the phonological properties of words, while taking word meaning into account. This is achieved by identifying the MBCDI words that the child understands, but does not say (words in the passive lexicon) and selecting words of higher PND for therapy sessions to expand the child's spoken lexicon. Vocabulary intervention to stimulate the production of words (e.g., modelled after Alt, Meyers, Oglivie, Nicholas & Arizmendi, 2014) should apply the principles of implicit learning, by manipulating input variability, input dose, and the phonological density of the treatment words. A follow-up review of the child's progress might indicate that further language intervention is required to ensure that the child increases the size of the active lexicon by beginning to develop phonological networks of sparser words.

Study Limitations

We identified four possible limitations. The first limitation is that we relied on parental report of a list of words and have not tested the accuracy of children's word productions. While the MBCDI is the instrument most often used in this type of research, this leads to a certain number of risks. One relates to the questionnaire: the list of words that is presented in the questionnaire is not exhaustive (Fenson et al., 1994). Furthermore, as noted above, we did not ask

Running head: COMPREHENSION-EXPRESSION GAP

parents what they intended when they completed the questionnaire, so we cannot be sure that parents distinguished between spontaneous speech and imitations (Bates, Dale & Thal, 1995). Another risk relates to those filling in the questionnaire: parents or caretakers could under- or over-report on their children's (language) skills (Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky, & Paradise, 2000; Houston-Price et al., 2007), which may be more of a risk for language comprehension than for production (Feldman et al., 2000; Westerlund & Sundelin, 2000). As noted earlier, reliability of parental scoring of individual words may be unstable (Yoder, Warren & Biggar, 1997), however, other research has suggested that parent scoring of word comprehension is accurate at the item level (Styles & Plunkett, 2009). Well-designed experiments are required to shed more light on this issue.

Additionally, reliability is stronger for the MBCDI toddler version (16-30 months), than the infant version (8-16 months). We have reported on data using the toddler version, for which reliability and validity has been reported in different studies, both for production (Feldman et al., 2005; Law & Roy 2005; Simonsen, Kristoffersen, Bleses, Wehberg & Jørgensen, 2014; Trudeau & Sutton, 2011) and comprehension (Simonsen et al., 2014). Note too that the Dutch CDI has been normed (Zink & Lejaegere, 2002) and has been reported to be sufficiently reliable (van Viersen et al., 2017). We take the NCDI to be a reliable and useful measure of vocabulary size. Nonetheless further empirical evidence and replication of existing findings of the validity of parent report at an item level would be valuable.

Reliance on this parental report measure also means that readers may be concerned that we did not test whether children's accuracy of spoken words, in addition to (sub)lexical characteristics, influences which words are in early spoken vocabularies. Some recent work has looked at questions related to this. Kehoe, Patrucco-Nachen, Friend and Zesiger (2018) examined the phonological and lexical characteristics of 40 French-speaking toddlers' lexicons using

Running head: COMPREHENSION-EXPRESSION GAP
the French version the MBCDI - L'Inventaire Français du Développement Communicatif (IFDC) (Kern & Gayraud, 2010). The children were 29 months old on average. The authors included measures of children's speech production accuracy in a spontaneous speech task. Their findings were consistent with our study: there was a relationship between PND and children's active vocabulary size, with PND accounting for 57% of the variance in vocabulary size. They also found a significant relationship between their measure of children's speech production accuracy (percentage of consonants correct) and vocabulary size ($r(38) = .62$) but this effect disappeared when the effect of PND was entered into a regression analysis. One other measure of phonological development, the number of consonants in children's syllable-initial phonetic inventories contributed an additional 8% of variance to the model once PND had been counted for. We can conclude that while children's phonological development may play a role in lexical development, PND is a stronger factor.

Further, Zamuner and Thiessen (2018) evaluated a variety of factors that could influence the likelihood that five English-speaking toddlers will imitate a new word. Words that were shorter in length and from denser neighborhoods were more likely to be spontaneously imitated. Additionally, words that contained sounds that the child has previously produced accurately were imitated more often.

A second limitation is that we only studied the predictors of monosyllabic words. This experimental control allowed for comparisons with previous studies and replication of findings across languages. However, limiting analyses to monosyllabic words may affect the external validity of the results. Future work should also include multisyllabic words, which typically have fewer phonological neighbors. The effect of neighborhood density on the full-sized lexicon might therefore be smaller, and other lexical factors (such as word frequency) may play a larger role.

Perhaps the most significant limitation is that the data provided a snapshot of a child's passive and active lexicons at one time point. Longitudinal MBCDI data are required to confirm our theory that PND is the driver of which words first transition from passive to active lexicons because of the ease of phonological encoding of denser words. Examination of longitudinal language samples would also provide confirmatory evidence of child word production (e.g., Van Viersen et al., 2017) and would allow for measurement of the phonetic accuracy of spoken words. It is possible that the size and composition of a child's phonetic inventory also predicts whether or not a known word is said.

Fourth, there are different measures for semantic neighborhood density and these should be compared in an analysis similar to that conducted here. In addition to other semantic factors - apart from semantic density - that may influence early word learning. For instance, nouns are usually acquired before verbs, which may reflect their higher imageability and concreteness (McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011). The earliest verbs tend to involve specific regions of the body, which can be explained in embodiment theory (Maouene, Hidaka, & Smith, 2008). Future word learning studies could compare all SND metrics as predictors of the probability of a known word being said, and may consider manipulating other semantic factors to further assess whether they play a role in the comprehension-expression gap. For instance, word grammatical categories should be examined separately (e.g., Kern & dos Santos, 2018).

Conclusion

In conclusion, we have demonstrated that the comprehension-expression gap may be accounted for by the phonological and word frequency characteristics of words, confirming Stokes's hypothesis of a phonological density advantage for spoken words. Importantly phonological neighborhood density operated similarly across children in this sample while there were

Running head: COMPREHENSION-EXPRESSION GAP

individual differences in the impact of word frequency. There are three major innovations in the current report - i) measurement of the passive lexicon, not only the active lexicon, ii) the incorporation of semantic neighborhood density as a possible predictor of whether or not a known word is said, and iii) the analysis of Dutch. The main contribution to our existing knowledge is that a phonological encoding gap may be a primary linguistic factor that accounts for the comprehension-expression gap in early word learning, over and above a semantic factor and independent of word frequency.

PRE-PROOF AUTHOR COPY

References

- Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically-based word learning intervention for late-talking toddlers. *Journal of Communication Disorders*, 52, 207-220. <https://doi.org/10.1016/j.jcomdis.2014.07.002>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical data base on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium. <https://catalog.ldc.upenn.edu/docs/LDC96L14/celex.readme.html>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and language*, 68(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv:1506.04967.v2* Downloaded from <https://arxiv.org/abs/1506.04967>
- Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language*, 96-151. Oxford: Basil Blackwell.
- Beckage, N., Aguilar, A., & Colunga, E. (2015). Modeling Lexical Acquisition Through Networks. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio, (Eds.) *Proceedings of the 37th Annual Meeting of the Cognitive*

Science Society, Austin: Cognitive Science Society, 190-195. URL: <http://www.proceedings.com/27265.html>

- Bloom, L. (1974). Talking, understanding, and thinking: Developmental relationship between receptive and expressive language. In R. Schiefelbusch & L. Lloyd (Eds.), *Language perspectives: Acquisition, retardation, and intervention*, 285–308. Baltimore, MD: University Park Press.
- Brybaert, M., Mandera, 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>
- Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, 75, 159-180. <https://doi.org/10.1016/j.jml.2014.05.005>
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205-15. <https://doi.org/10.1017/S0305000900013180>
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighborhoods in the developing lexicon. *Journal of Child Language*, 30, 441-469. <https://doi.org/10.1017/S030500090300557>
- Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering* 14(3), 391–403. <https://doi.org/10.1081/QEN-120001878>
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*, pp. 125-127. <https://doi.org/10.3758/BF03203646>

- Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*, 544-560.
doi:10.1044/1092-4388(2003/044)
- De Bree, E., Zamuner, T.S., & Wijnen, F. (2014). Neighborhood densities in the vocabularies of Dutch children with a familial risk of dyslexia. In R. Kager, J. Grijzenhout, & K. Se-bregts (Eds.), *Where the Principles Fail. A Festschrift for Wim Zonneveld on the Occasion of his 64th Birthday*, 17-28. Utrecht: Utrecht institute of Linguistics OTS. <https://dspace.library.uu.nl/handle/1874/307068>
- De Deyne, S., Navarro, D., Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods, 45* (2), 480-498. <https://doi.org/10.3758/s13428-012-0260-7>
- De Deyne S., & Storms G. (2008a). Word Associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*,198-205. <https://doi.org/10.3758/BRM.40.1.198>
- Evert, S. (2008). Corpora and collocations. In A. Ludeling & M. Kytö (Eds.). *Corpus Linguistics: An international handbook*, 1-53. URL <https://www.degruyter.com/view/serial/39720>
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two Years. *Child Development, 71*, 310–322. <https://www.jstor.org/stable/1131991>
- Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child

Running head: COMPREHENSION-EXPRESSION GAP
language at ages 2 and 3 years. *Child Development*, 76(4), 856-68. <https://doi.org/10.1111/j.1467-8624.2005.00882.x>

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

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5). <http://dx.doi.org/10.2307/1166093>

Frank, M. C., Braginsky, M., Yurovsky, & Marchman, V. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, <https://doi.org/10.1017/S0305000916000209>

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, 114(4), 489-509. <https://doi.org/10.1016/j.jecp.2012.11.005>

Gordon, J. K., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language*, 79, 21-31. <https://doi.org/10.1006/brln.2001.2574>

Gordon, K. R. (2019). How mixed-effects modeling can advance our understanding of learning and memory and improve clinical and educational practice, *Journal of Speech, Language, and Hearing Research*, 62, 507-524. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0240

Gordon, K. R., McGregor, K. K., Waldier, B., Curran, M.K., Gomez, R.L. & Samuelson, L.K.

(2016). Preschool children's memory for word forms remains stable over several days, but gradually decreases after 6 months. *Frontiers in Psychology*, 7:1439.

<https://doi.org/10.3389/fpsyg.2016.0143>

Hamilton, A., Plunkett, K., & Schafer, G., (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower scores in the UK than the USA. *Journal of Child Language*, 27, 689-705. <http://dx.doi.org/10.1017/S0305000900004414>

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks. *Psychology Science*, 20(6), 729–739.

<https://doi.org/10.1111/j.1467-9280.2009.02365.x>

Horst, J. S. and Samuelson, L. K. (2008), Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13: 128-157. <https://doi.org/10.1080/15250000701795598>

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, 34, 701-724. <https://doi.org/10.1017/S0305000907008124>

Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, 53, 258-276. <https://doi.org/10.1016/j.jml.2005.03.002>

Jimenez, E. & Hills, T. (2017). *Network Analysis of a Large Sample of Typical and Late Talkers*, Paper presented at the Annual Conference of the Cognitive Science Society. URL <https://mindmodeling.org/cogsci2017/papers/0438/paper0438.pdf>

- Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *PNAS*, *96*(13), 7592-7597. <https://doi.org/10.1073/pnas.96.13.7592>
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944-946. <https://doi.org/10.1111/2041-210X.12225>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *1*(13). <https://doi.org/10.18637/jss.v082.i13>
- Kehoe, M., Patrucco-Nanchen, T., Friend, M., & Zesiger, P. (2018). The relation between phonological and lexical development in French-speaking children. *Clinical Linguistics & Phonetics*. <https://doi.org/10.1080/02699206.2018.1510984>
- Kerkhoff, A., de Bree, E., de Klerk, M., & Wijnen, F. (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, *40*, 11–28. <https://doi.org/10.1017/S0305000912000098>
- Kern, S., & dos Santos, C. (2018). Invariance in variation: Frequency and neighborhood density as predictors of vocabulary size. In M. Hickmann, E. Veneziano, J. J. (Eds). *Sources of variation in first language acquisition: Languages, contexts, and learners*, 183-200. Amsterdam: John Benjamins Publishing Company.
- Kern, S., & Gayraud, F. (2010). *L'inventaire français du développement communicatif*. Grenoble, France: Editions La Cigale. URL: <https://afpa.org/outil/inventaires-francais-developpement-communicatif/>

- Law, J., Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health, 13*, 198–206. <https://doi.org/10.1111/j.1475-3588.2008.00503.x>
- Li, Y., Bandar, Z. A., & McLean, D. A. (2003) An approach for measuring semantic similarity between words using multiple information sources. *IEEE transactions on knowledge and data engineering, 15 (4)*, 871-882. <https://doi.org/10.1109/TKDE.2003.1209005>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49(4)*, 1494-1502. <https://doi.org/10.3758/s13428-016-0809-y>
- Majorano, M., Vihman, M. M. & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10(2)*, 179-204. <https://doi.org/10.1080/15475441.2013.829740>
- Maekawa, J., & Storkel, H. L. (2006). Individual differences in the influence of phonological characteristics on expressive vocabulary development by young children. *Journal of Child Language, 33(03)*, 439-459. <https://doi.org/10.1017/S0305000906007458>
- Maouene, J., Hidaka, S., & Smith, L. B. (2008). Body parts and early-learned verbs. *Cognitive Science, 32(7)*, 1200-1216. <https://doi.org/10.1080/03640210802019997>
- Marinus, E., Nation, K., & de Jong, P. F. (2015). Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch. *Journal of Experimental Child Psychology, 139*, 127-147. <https://doi.org/10.1016/j.jecp.2015.05.006>
- McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word

Running head: COMPREHENSION-EXPRESSION GAP
learning. *Developmental Science*, 14(2), 181-189. [https://doi.org/10.1111/j.](https://doi.org/10.1111/j.1467-7687.2010.00968.x)

1467-7687.2010.00968.x

- Menn, L., & Matthei, E. (1992). The “two-lexicon” account of child phonology: looking back, looking ahead. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, Implications*, 211–247. Timonium, MD: York Press.
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3(41), <https://doi.org/10.3389/fpsyg.2012.00041>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134). [doi:10.1098/rsif.2017.0213](https://doi.org/10.1098/rsif.2017.0213)
- Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402–407. <https://doi.org/10.3758/bf03195588>
- Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, 119(3), 517. <https://doi.org/10.1037/a0028450>
- Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. USA: Springer.
- Sahni, S. D., & Rogers, T. T. (2008). Sound versus meaning: What matters most in early word learning? *Proceedings of the 7th IEEE International Conference Monterey, CA*. <https://doi.org/10.1109/DEVLRN.2008.4640843>
- Schneider, R. M., Yurovsky, D., & Frank, M. C. (2015). Large-scale investigations of

Running head: COMPREHENSION-EXPRESSION GAP
variability in children's first words. *Proceedings of the 37th Annual Conference of the
Cognitive Science Society*. URL [https://langcog.stanford.edu/papers/SYF_underr-
view.pdf](https://langcog.stanford.edu/papers/SYF_underreview.pdf)

- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*(1), 3–23. <https://doi.org/10.1177/0142723713510997>
- Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research*, *53*, 670-683. [https://doi.org/10.1044/1092-4388\(2009/08-0254\)](https://doi.org/10.1044/1092-4388(2009/08-0254))
- Stokes, S. F. (2014). The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language*, *41*(3), 634-657. <https://doi.org/10.1017/S030500091300010X>
- Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry* *50*(4), 498-505. <https://doi.org/10.1111/j.1469-7610.2008.01991.x>
- Stokes, S.F., Bleses, D., Basbøll, & Lambertsen, C. (2012a). Statistical learning in emerging lexicons: The case of Danish. *Journal of Speech, Language, and Hearing Research*, *55*, 1265- 1273. [https://doi.org/10.1044/1092-4388\(2012/10-0291\)](https://doi.org/10.1044/1092-4388(2012/10-0291))
- Stokes, S. F., Kern, S., & dos Santos, C. (2012b). Extended statistical learning as an account for slow vocabulary growth. *Journal of Child Language*, *39*, 105-129. <https://doi.org/10.1017/S0305000911000031>

- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language & Hearing Research, 44*, 1321–1337. [https://doi.org/10.1044/1092-4388\(2001/103\)](https://doi.org/10.1044/1092-4388(2001/103))
- Storkel, H. L. (2004a). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics 25*, 201-221. <https://doi.org/10.1017/S0142716404001109>
- Storkel, H. L. (2004b). Methods for minimising the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research, 47*, 1454-1468. [https://doi.org/10.1044/1092-4388\(2004/108\)](https://doi.org/10.1044/1092-4388(2004/108))
- 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. <https://doi.org/10.1017/S030500090800891X>
- Styles, S. J., & Plunkett, K. (2009). What is ‘word understanding’ for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language, 36*, 895-908. <https://doi.org/10.1017/S0305000908009264>
- Takáč, M., Knott, A., & Stokes, S. F. (2017). What can neighborhood density effects tell us about word learning? Insights from a connectionist model of vocabulary development. *Journal of Child Language, 44*, 346-379. <https://doi.org/10.1017/S0305000916000052>
- Trudeau, N., Sutton, A. (2011). Expressive vocabulary and early grammar of 16- to 30-month-old children acquiring Quebec French. *First Language, 31*, 480–507. <https://doi.org/10.1177/0142723711410828>

van de Weijer, J. (1998). *Language input for word discovery* (Doctoral dissertation). Max

Planck Series in Psycholinguistics 9. URL https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2057670_5

van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de

Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of Dyslexia. *Journal of Speech, Language, and Hearing Research*, 60(4), 937-949.

https://doi.org/10.1044/2016_JSLHR-L-16-0031

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech

production. *Journal of Psychological Learning Memory and Cognition*, 28(4), 735-747.

<https://doi.org/10.1037/0278-7393.28.4.735>

Westerlund, M., & Sundelin, C. (2000). Screening for developmental language disability in 3-year-old children. Experiences from a field study in a Swedish municipality. *Child: Care, Health & Development*, 26, 91 – 110. <https://doi.org/10.1046/j.1365-2214.2000.00171.x>

Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, 6(1), 59-64. <https://doi.org/10.1044/1058-0360.0601.59>

Zamuner, T. S., & Thiessen, A., (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics*, Published online 5th April. <https://doi.org/10.1017/cnj.2018.10>

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29. <http://dx.doi.org/10.1037/0033-2909.131.1.3>

Running head: COMPREHENSION-EXPRESSION GAP

Zink, I., & Lejaegere, M. (2002). *N-CDIs: Lijsten voor Communicatieve Ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al.* [A CDI user's manual with normative and validity data]. Leusden, The Netherlands: Acco. URL <https://www.acco.be/nl-be/items/9789033448997/N-CDIs--Lijsten-voor-Communicatieve-Ontwikkeling>.

PRE-PROOF AUTHOR COPY

Table 1

Summary of the Mixed Effects Model for the Fixed Effects of WF, PND and SND, Their Interactions, and the Random Effects of Subject and Item.

	SE	Estimate	Z	p	CI	
Fixed effects						
Intercept	0.22	2.02	9.17	0.001	1.59	2.46
WF	0.10	0.29	2.95	0.003	0.09	0.49
PND	0.09	0.27	2.77	0.005	0.07	0.46
SND	0.09	-0.04	-0.43	0.66	-0.23	0.15
WF:PND	0.09	0.09	1.03	0.30	-0.08	0.28
WF:SND	0.10	-0.15	-1.42	0.15	-0.37	0.05
PND:SND	0.08	-0.05	-0.62	0.53	-0.22	0.11
WF:PND:SND	0.10	0.02	0.27	0.78	-0.17	0.23
Random effects			Variance	SD		
Item (Intercept)			2.63	1.62		
Subject(Intercept)			7.79	2.79		
Subject:WF (Slope)			0.04	0.21		

Note. WF = word frequency, PND = phonological neighborhood density, SND = semantic neighborhood density, SE = standard error, CI = Confidence Interval (95%).

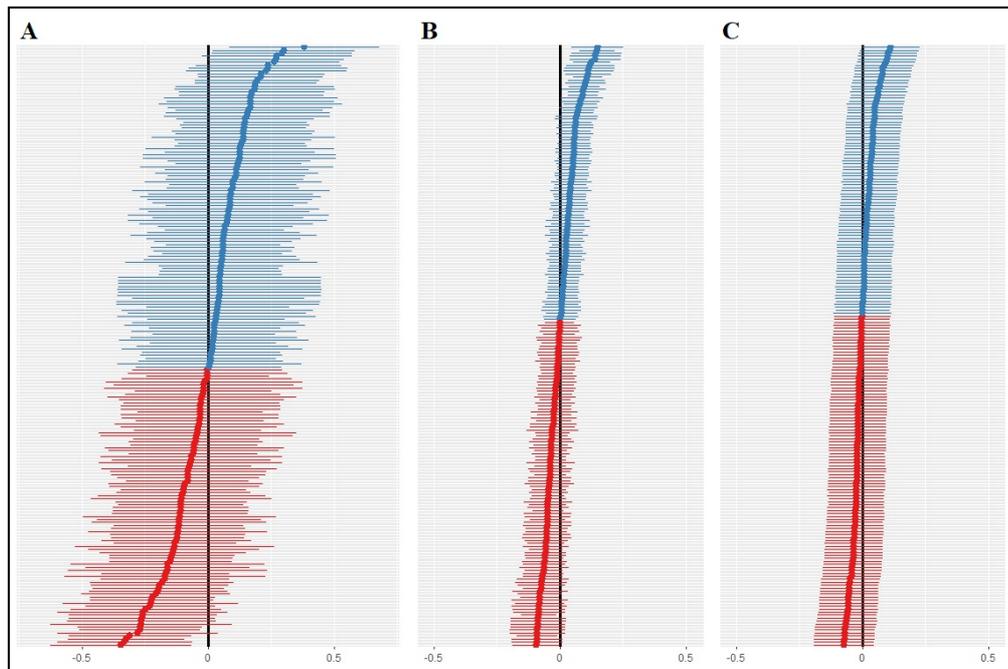


Figure 1. Random slopes for WF(A), PND(B) and SND(C) across all children. The Y axis represents every individual child on one horizontal line, the X axis represents Log Odds with 0 showing the intercept or midpoint. Dots are posterior means and the shading indicates the 95% confidence intervals. The blue lines indicate the children in whom WF, PND, and SND had higher effects compared with the general effect or intercept, and the red lines indicate those with lower effects or coefficients than the intercept.

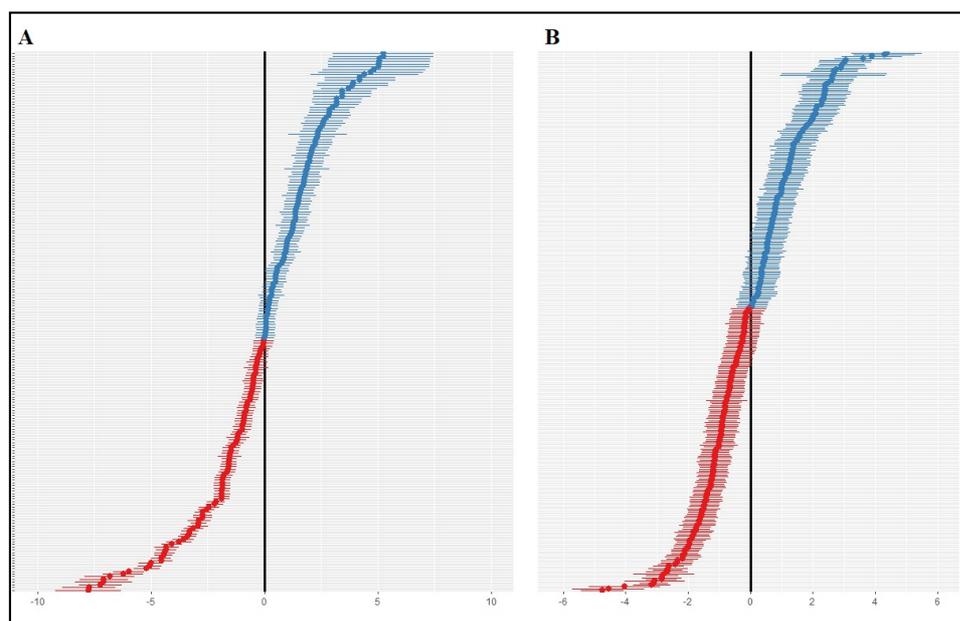


Figure 2. Random intercepts for individual children (A) and items (B). The X axis in both A and B represents Log Odds with 0 being the intercept or midline. In A the Y axis represents children, in B the Y axis represents items. Dots are posterior means and the shading indicates 95% confidence intervals. In A the blue and red lines indicate the children with larger and smaller active lexicons than the average, respectively. In B the blue and red lines indicate the words which appeared more and less in the children's active lexicon on average, respectively.

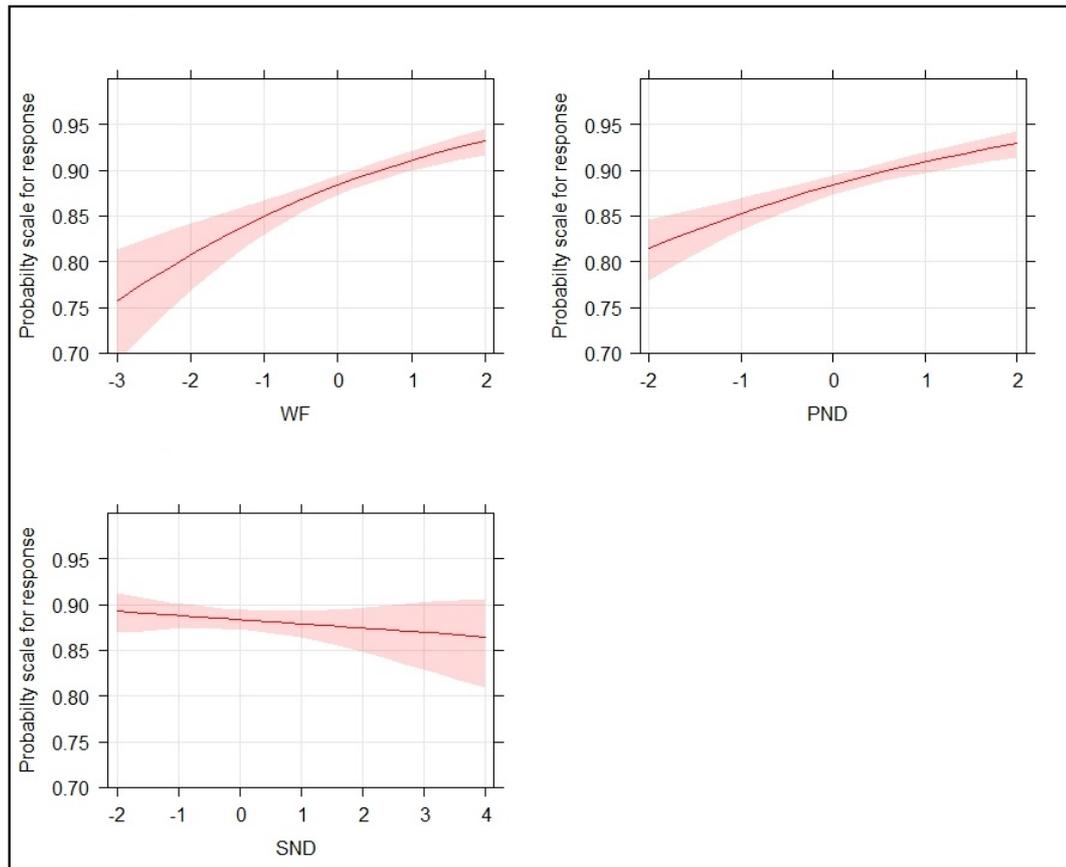


Figure 3. Slope of WF, PND, and SND scaled values with 95% confidence interval. The Y axis represents the probability of the parent response = child ‘says word’, and the X axis represents standardized values.