CrossMark

Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth

Gary Jones [a,*], Caroline F. Rowland [b,c]

[a] Nottingham Trent University, UK
[b] Max Planck Institute for Psycholinguistics, Netherlands
[c] University of Liverpool, UK

ARTICLE INFO

ABSTRACT

Children who hear large amounts of diverse speech learn language more quickly than children who do not. However, high correlations between the amount and the diversity of the input in speech samples makes it difficult to isolate the influence of each. We overcame this problem by controlling the input to a computational model so that amount of exposure to linguistic input (quantity) and the quality of that input (lexical diversity) were independently manipulated. Sublexical, lexical, and multi-word knowledge were charted across development (Study 1), showing that while input quantity may be important early in learning, lexical diversity is ultimately more crucial, a prediction confirmed against children's data (Study 2). The model trained on a lexically diverse input also performed better on nonword repetition and sentence recall tests (Study 3) and was quicker to learn new words over time (Study 4). A language input that is rich in lexical diversity outperforms equivalent richness in quantity for learned sublexical and lexical knowledge, for well-established language tests, and for acquiring words that have never been encountered before.

1. Introduction

The onset and speed of vocabulary acquisition differs enormously from child to child, in all languages. A significant proportion of this variation is explained by children's different experiences of language in their environment. Children who hear a large amount of lexically diverse child-directed speech learn language more quickly and have larger vocabularies than children who do not (for a review, see e.g. Hoff, 2006). However, it is almost impossible to directly compare the effect of input quantity with the effect of lexical diversity in natural speech samples because the two are highly correlated. In this paper, we use computational modeling to investigate how the quantity of input and the lexical diversity of that input impact on a model's learning throughout development. We show how the lexical diversity of the language input is more important than its quantity for learning sublexical and lexical knowledge, for performing language tests that are highly predictive of the children's language ability, and for vocabulary learning more generally.

* Corresponding author at: Department of Psychology, Nottingham Trent University, Chaucer Building, Nottingham NG1 4FQ, UK.
E-mail address: gary.jones@ntu.ac.uk (G. Jones).

Preschool children show large individual differences both in the age of onset, and in the speed, of vocabulary learning (Fenson et al., 2007). These differences are strongly predicted by environmental factors, particularly the quantity and quality of the linguistic input children receive. In terms of quantity, several studies have established a direct strong or medium correlation between measures of the sheer quantity of linguistic input children hear and the size of children's vocabularies (see e.g. Bornstein, Haynes, & Painter, 1998; Bornstein & Tamis-LeMonda, 1995; Cartmill et al., 2013; Hart & Risley, 1992; Hart & Risley, 1995; Hoff & Naigles, 2002; Hurtado, Marchman, & Fernald, 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2012). For example, Huttenlocher et al. (1991) showed that 20% of the variance in children's vocabulary growth between 16 and 24 months of age (N = 22) was predicted by variations in the amount of speech produced by the children's mothers. Analyses of Hart and Risley's (1992) and Hart and Risley's (1995) American data show that some children (particularly those defined as of high socio-economic status) were exposed to up to 153,000 more words per week than others, which had a significant effect on these children's vocabulary growth. Pearson, Fernandez, Lewedeg, and Oller (1997) even showed that the relative vocabulary size of bilingual English-Spanish speaking one-to-two year olds was predicted by the relative amount of input they received in each language, which suggests a direct correlation between the likelihood of hearing a word in a language and the ability to learn it (for similar results, see also Hoff et al., 2012).

The quality, as well as the quantity, of the language input is important. Input quality can be operationalized in many different ways. Some studies have tested the complexity of input utterances, using for example, a Mean Length of Utterance (MLU) measure (Bornstein et al., 1998; Hoff & Naigles, 2002) or constituent and clausal complexity (Huttenlocher et al., 2010). Others have focused on communicative devices thought to increase the likelihood of the child interpreting a word correctly, such as the degree to which the referent of the word is easily identifiable or not (referential uncertainty; Cartmill et al., 2013) or the number of utterances spoken during periods of joint attention (Hoff & Naigles, 2002). Still others focus on features of the conversations that parents hold with their children, and measure, for example, how effectively parents engage the child's attention during episodes of joint engagement or how often they model language during routines and rituals (e.g. Hirsh-Pasek et al., 2015). However, the metric that is used most often, and which produces the most consistent effects, is lexical diversity, operationalized as the number of different word types produced by the caregiver during a set timeframe (e.g. Bornstein et al., 1998; Demir-Vegter, Aarts, & Kurvers, 2014; Hoff & Naigles, 2002; Huttenlocher et al., 2010; Pan, Rowe, Singer, & Snow, 2005; Rowe, 2012; see also Hsu, Hadley, & Rispoli, 2015, for results for verb lexical diversity). For example, Bornstein et al. (1998) found that both maternal lexical diversity and maternal MLU were significant predictors of child vocabulary at 18 months in a structural equation model. Similarly, Hoff and Naigles (2002) reported that the number of different word types produced by the mothers of 63 two-year olds was a strong predictor of the number of different words their children produced ten weeks later (though in this study, MLU was a stronger predictor). An input that models a high proportion of rare words is particularly effective, especially at older ages (e.g. Beals, 1997; Rowe, 2012; Weizman & Snow, 2001). Thus, there is good evidence for a role for both the quantity of the input and its lexical diversity in determining the rate at which children develop vocabulary.

However, almost all research to date has tested input quantity and lexical diversity using speech samples where both are free to vary. Very few studies, if any, directly compare the independent effects of input quantity and lexical diversity on child language, and virtually none test how these relationships change with development. This is problematic, since any complete theoretical account of vocabulary acquisition has to explain not only how different properties of the input are utilized during learning, but also how changes in the amount and structure of this learned knowledge affect subsequent learning.

Studying the effect of the input developmentally is important since there is tentative evidence for differential age-related effects. Rowe (2012) recently assessed age-related changes in the contribution of input quantity and quality to vocabulary growth, looking at the effects of input at 18, 30, and 42 months on children's vocabulary 12 months later. Quantity (number of word tokens produced in a 90 min interaction) was a strong predictor of vocabulary at 30 months but lexical diversity (total number of different word types) and sophistication (total number of rare words) were the strongest predictors by the time the children reached 42 months. By 54 months, the amount of decontextualized talk in the input (explanations, talk about pretense, narratives) was the strongest predictor of vocabulary. This progression makes sense when seen in the light of the acquisition tasks facing children at different ages. In the earliest stages of learning, when the child knows few words, hearing any word multiple times will lead to learning. Thus, frequent exposure to a small number of words will be effective. However, later on, when many high frequency words have already been learnt, exposure to a larger variety of words becomes more important. Even later, decontextualized language introduces children to a different set of (infrequent) words (e.g. explicit explanation of word meanings, words about the future or past), which is why it has its strongest effect later in learning.

In this paper, we directly compare the effects of input quantity and lexical diversity on vocabulary acquisition across development. We do this using a computational model, because directly comparing these effects is very difficult to do within the context of natural language settings. Parents who are verbose also use a wider variety of different words (e.g., Hart & Risley, 1992; Smolak & Weinraub, 1983; Weizman & Snow, 2001), so, while maternal type and token data might both predict child vocabulary growth, the correlation between them is usually so high as to render meaningless any statistical differences between the two. For example, Hoff and Naigles (2002) were unable to enter both number of word tokens (input quantity) and number of word types (lexical diversity) into the same regression model because high correlations between these measures ($r = 0.89$) made their effects indistinguishable statistically. The majority of research outlined above is therefore difficult

to interpret since measures of input quantity are confounded by measures of input quality (lexical diversity, syntactic complexity etc.) and vice versa.

Using a computational model allows us to bypass this problem, because we can artificially manipulate the naturalistic language input by varying the amount of language input (while keeping lexical diversity fixed) or varying its lexical diversity (while keeping amount of input fixed). By measuring the model's learning after it has been trained using different inputs we can isolate the different effects. In addition, by testing the model at different points in its learning cycle, and comparing its output with data from real children, we can investigate whether the relative importance of the quantity and diversity of the input changes throughout development and how the model's developing vocabulary impacts upon its performance.

In computational modeling, much hinges on the type of learning mechanisms and parameters implemented in the model. If the results of the model are to inform our thinking about how children might learn, these mechanisms must be psychologically plausible. To this end, we chose to use CLASSIC (Jones, 2016; Jones, Gobet, Freudenthal, Watson, & Pine, 2014), a computational model that learns lexical and phonological knowledge from exposure to word-delimited phonemically-coded input, and has been successful previously in modeling children's performance in closely related developmental tasks (nonword repetition performance, see e.g. Jones, Gobet, & Pine, 2007). In addition, since CLASSIC can use naturalistic (maternal) utterances as input, we can manipulate the properties of this input (e.g., increasing quantity or lexical diversity) to examine the effect this has on the model's learned knowledge, its performance on established language tests, and also how the model's vocabulary learning compares with that of real children. Furthermore, since CLASSIC learns incrementally from the input, we can access the results of the model's learning at different points in development.

In sum, the goal of this paper was to determine the effect of the quantity and diversity of the input on vocabulary learning in a CLASSIC model. In Study 1, we tested the effect of input quantity and lexical diversity on the model's learning. In Study 2, we compared the model's vocabulary learning with that of 2- to 3-year-old children. In Study 3, we investigated the effect of input quantity and diversity more generally by examining the performance of the model on two tests that are highly predictive of children's language success: nonword repetition and sentence recall. In Study 4, we examined the effect of different inputs on the process of word learning itself, by determining how many new words were learned by the models exposed to different input types. Below we first outline the model and its parameters, before presenting the four studies and discussing their implications.

2. CLASSIC (Chunking lexical and sublexical sequences in children)

2.1. How CLASSIC learns

CLASSIC is a symbolic model based on the EPAM/CHREST architecture, which has been used to model a range of perceptual and cognitive phenomena in verbal learning, expert memory, language acquisition and education (see Gobet et al., 2001). These models implement a chunk-based learning mechanism that learns by gradually chunking information in the model's internal representational system on the basis of incoming input. CLASSIC's architecture has been adapted to learn phonemic sequence vocabulary from phonetically-coded naturalistic input.

CLASSIC begins with built-in knowledge of the phonemes of standard British English, realized as text symbols that are analogous to the International Phonetic Alphabet. The input to the model is real utterances (e.g., maternal utterances) that are converted into text-based word-delimited phonetic symbols (e.g., /w ɒ t/ k æ t/ for 'what cat?'). From this input, CLASSIC learns increasingly long sequences (chunks) of phonetic information. The initial presentations of phonetic input will usually lead to sublexical information being learned (i.e., phonetic sequences below the lexical level). However, as more input is seen, it is increasingly likely that the sequences learned will correspond to lexical items and multi-word sequences.

For the purposes of this article, we will refer to the knowledge represented in CLASSIC as chunks (we will see that a chunk can contain a phoneme, a sublexical sequence of phonemes, a word or sequence of words). Take as an example the utterance /w ɒ t/ i z /θ æ t/ ('what is that?'). CLASSIC uses its existing knowledge to code the utterance into as few chunks as possible. Since CLASSIC begins with chunks that correspond to each phoneme within standard British English, on its first presentation, the utterance will be coded into eight chunks, one for each phoneme (/w, ɒ, t/ i, z /θ, æ, t/). CLASSIC's learning algorithm will then create new chunks that combine the knowledge from adjacent chunks (i.e., wɒ, ɒt, iz, θæ, æt). CLASSIC does not learn across word boundaries unless the chunks involved correspond to lexical items or multi-word sequences (for an explanation, see below). If the same utterance were presented a second time, it would now be coded using five chunks (/wɒ, t/ iz, /θæ, t/) and the learning algorithm would again learn new chunks for adjacent items (i.e., wɒt, θæt). On the third presentation, the utterance can be coded into chunks that correspond to lexical items (/wɒt/, iz, /θæt/), which would lead to multi-word chunks being learned (i.e., wɒt iz, iz θæt). Within CLASSIC, chunks are represented in a hierarchy based on the initial phoneme in a learned sequence (e.g., below /w/ would be the sublexical chunk /wɒ/, below that the lexical chunk /wɒt/, and below that the multi-word chunk /wɒt iz/). However, the hierarchical nature of CLASSIC is not used in any of the results presented here – rather it is the presence or absence of learned chunks that is important. Knowledge of whether a newly learned chunk is sublexical, lexical or multi-word is determined based on the word-delimited input.

Young children are usually unable to process long utterances in full (e.g., Gathercole, 2006), and so CLASSIC includes a processing constraint that limits the amount of information that can be processed at any one time. It is based on a simple sigmoidal function (see Fig. 1) in which, on average, 4.5 chunks are processed for any given input, with the most recently
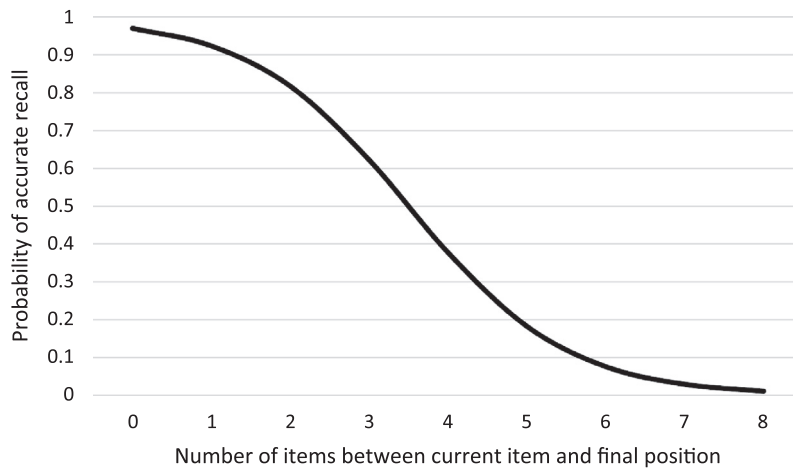
Fig. 1. Probability of accessing a chunk on the basis of its position from the end of the input utterance. '0' on the x-axis represents a chunk in the utterance-final position, '1' represents the last but one chunk and so on. A probability of 0.5 occurs 4.5 chunks from the end of the utterance (at 3.5 on the x-axis because the final item is indicated by '0').

heard chunks being favored. We chose a sigmoidal function because this is broadly consistent with the effects found in serial recall tests, where the proportion of lists recalled is a sigmoidal function of list length (see Grenfell-Essam & Ward, 2012). For example, in Fig. 1, the two most recent chunks have access probabilities of 0.97 and 0.92, meaning the probability of their joint access is 0.89 (0.97 ∗ 0.92 = 0.89). The processing limitation restricts learning from longer words and utterances because it means that consecutive chunks are not always accessed and, therefore, not always chunked. For example, assuming CLASSIC hears a single-word utterance, where the word can be coded using 9 chunks, and where 4.5 chunks will be accessed on average, the probabilities outlined in Fig. 1 mean that 2.5 adjacent chunks will be accessed (i.e. the model has 2.5 opportunities to learn something about the word). Table 1 exemplifies how learning is reduced on the basis of the processing constraint. It is worth noting that previous applications of CLASSIC have labeled the constraint on information processing as 'short-term memory'; however, since a limit on information processing could arise from many sources (e.g., limits on attention), we now use the broader term processing constraint.

As Table 1 illustrates, on each presentation of the utterance, CLASSIC learns more information (e.g., longer chunks). As a consequence, it is capable of processing more of the utterance on the next presentation, which leads to further learning in turn. In other words, early in the learning cycle, the model is unable to represent, and learn from, much of the input. Later in learning, more of the input can be processed, and more learning is likely to take place.

The more input the model receives, the more vocabulary it learns (chunks that correspond to a lexical item), and the better its performance on vocabulary-related tasks. However, the quantity and lexical diversity of the input are likely to have different effects on the model's learning. Increasing the quantity of the input (number of word tokens) while keeping lexical diversity (number of word types) constant is likely to lead to the model learning increasingly longer chunked units of consistently repeated sequences of phonemes. Increasing the lexical diversity of the input, on the other hand, is likely to lead to it learning a wide variety of smaller chunks, because it is less likely that the same word, utterance or sequence of words will be repeated in the input. The question we ask is whether learning longer chunks of knowledge (from increased quantity) will lead to faster vocabulary learning and better performance on language-related tests than learning smaller but more varied chunks of knowledge (from increased lexical diversity). Increasing input quantity may lead to more words being learnt, or may simply lead to more chunking of words into multi-word chunks. Increasing lexical diversity may cause the model to learn more lexical chunks (words), or more sublexical chunks, with fewer of these chunked into words.

Table 1
Potential learning when CLASSIC is repeatedly exposed to the phrase 'what is that' when the input utterance is constrained by processing limitations. Items underlined are those that are not accessed based on the sigmoidal function from Fig. 1.

Number of exposures	Chunks to be subjected to processing constraint	New chunks learned
1. /w ɒ t/ i z /θ æ t/	/w̲, ɒ̲, t̲/ i̲, z /θ, æ, t/ [8 chunks, 50% accessed]	θæ, æt
2. /w ɒ t/ i z /θ æ t/	/w̲, ɒ̲, t/ i, z /θæ, t/ [7 chunks, 71% accessed]	iz, θæt
3. /w ɒ t/ i z /θ æ t/	/w̲, ɒ, t/ iz /θæt/ [5 chunks, 80% accessed]	ɒt, iz θæt

We investigated this in Study 1. However, before that, we justify the assumptions and parameters used within CLASSIC and then give details on how CLASSIC has been applied to verbal tasks in order to illustrate how the model can be seen as a viable model of language learning.

2.2. CLASSIC's assumptions and parameters

2.2.1. CLASSIC's learning mechanism

Like other models based on the EPAM/CHREST architecture (Gobet et al., 2001), CLASSIC implements a chunk-based learning mechanism that makes gradual changes to the model's internal representations by chunking information received from the incoming input. Chunking as an information-processing mechanism was first proposed by De Groot (1978), based on studies of human problem solving, and by Miller (1956), based on studies of human perception and memory. Since then, chunking has become established as one of the key mechanisms of human cognition (Gobet et al., 2001), and explains how humans are able to extract, and process, increasingly large amounts of information from the environment over time in the face of constant cognitive limitations. For example, by chunking information in long-term memory, we can encode and retrieve long sequences of numbers (e.g., in the digit span task, Richman, Staszewski, & Simon, 1995) and syllables (e.g., in nonword repetition tasks, Jones et al., 2007). Chunk-based learning models based on the EPAM/CHREST architecture have simulated human behavior in a variety of domains including verbal learning, expert memory for chess positions, science education, and language development (see Gobet et al., 2001, for a review).

CLASSIC's learning is based on two assumptions: (a) that regularly occurring sequences (be they sequences of phonemes or words) will eventually be learned as chunks; and (b) that the processing of a sequence is facilitated when the sequence can be coded using few, rather than many, chunks (see Table 1). Sublexically, these assumptions are based on robust evidence that children and adults find it easier to process and produce a phoneme sequence if they have previously been exposed to it as a sequence, than if they have simply been exposed to its component phonemes in separate phonological contexts (e.g., Catts & Kamhi, 1984; Demuth, Culbertson, & Alter, 2006; Jakobson, 1941/1968). For example, even when children are able to produce all constituent sounds in a phoneme sequence, they are more accurate at producing those sequences that exist in the native language over those that do not (Beckman & Edwards, 1999). There is also support at the lexical level: young children more readily process highly familiar sequences of words than less familiar ones (Bannard & Matthews, 2008) while multi-word sequences show age of acquisition processing advantages in the same way that lexical items do (e.g., faster response times to early acquired phrases, Arnon, McCauley, & Christiansen, 2017).

2.2.2. Parameter 1: Probability of learning a new chunk set to 1.0

CLASSIC learns new chunks by joining the information in two adjacent chunks together. This happens every time two adjacent chunks are accessed, which means that the probability of learning a new chunk when adjacent chunks are accessed is 1.0 (note that the processing constraint means that adjacent chunks are not always accessed). However, this fast learning is not intended to be an illustration of one-shot learning. Rather, it is a necessity based on the paucity of naturalistic language that is available as input to the model. The maternal input to the model (see later) averages 97,000 words, which is trivial compared to estimates of children's linguistic exposure. Hart and Risley (1992) and Hart and Risley (1995) have shown that some children are exposed to up to 153,000 more words per week than others, while Swingley (2007) suggests that children may be hearing up to half a million words in a three-week period. Thus, in order to see any development in our model, we have to allow it to learn more quickly than real children. That said, while our main model uses a chunk-learning probability set to 1.0, we also manipulated this in Study 1, to examine how different learning parameters behave in response to changes in input quantity and lexical diversity.

2.2.3. Parameter 2: Processing limitation of 4.5 chunks

A probabilistic processing constraint is applied within CLASSIC such that, on average, only 4.5 chunks can be accessed in any given input. This mimics processing limitations that prevent young children from processing long utterances in full (Gathercole, 2006). However, although we think it important to model some type of information processing constraint, we do not intend to imply that a setting of exactly 4.5 chunks is realized in young children. In Study 1, we modeled variations in the processing limitation, and show that the same broad pattern of effects emerges regardless of the number of chunks that can be processed at any one time.

2.2.4. Favoring the most recent chunks in an utterance (recency effect)

We also set the processing constraint to favor the most recent chunks in a coded utterance. This is based on well-established recency effects in which participants are more accurate at recalling the most recent (last presented) items in a list than items in the middle of lists (e.g., Gupta, 2003; Hulme et al., 1997). Young children also show a recency effect in that they attend more closely to word-final syllables (Echols & Newport, 1992). However, in Study 1, we also tested the effect of removing the recency bias. We do not implement a primacy effect (i.e., facilitatory recall for items at the beginning of to-be-recalled lists) because such effects are likely to result from a rehearsal strategy, which children below the age of 7 years rarely use (e.g., Gathercole & Adams, 1994).

2.2.5. Inclusion of word boundary information

We include information about word boundaries in the input because there is substantial evidence that children are already capable of determining word boundaries via a range of phonetic, phonological and distributional cues by their first birthday (see Rowland, 2014, for a review). However, it is worth noting that previous studies that either use a derivative of CLASSIC that dispensed with word boundaries, or that directly compare models with/without word boundary information, show that the inclusion/exclusion of word boundary information has very little effect on the performance of the model (Jones, 2016; Jones et al., 2007). For example, in Jones (2016), a direct comparison of models that included/excluded word boundaries in CLASSIC only moved the fit between model and child nonword repetition data from 7% to 8%.

2.3. Previous applications of CLASSIC to verbal tasks

The main strength of CLASSIC is how well it is able to simulate child and adult performance on tasks that are verbal in nature. Nonword repetition (NWR) is one such task, being highly predictive of vocabulary size (e.g., Gathercole & Baddeley, 1989; Hoff, Core, & Bridges, 2008), and of language performance more generally (e.g., Gathercole, 2006). NWR performance can even be used as a clinical marker of language impairment (e.g., Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Nash & Donaldson, 2005). As such, it provides an excellent test bed for CLASSIC's ability to approximate children's language learning.

For example, Jones (2016) examined the development of nonword repetition performance using the same CLASSIC model outlined here. The author analyzed the model's performance using input aimed at young children (maternal input addressed to 2–3-year-olds) and input addressed to older children (4–6 year olds). The linguistic input was temporally ordered so that utterances aimed at younger children were processed first. In this way, model performance at an early stage of learning could be compared to two-year-old children, model performance after more learning could be compared to three year olds, and so on until all of the input had been processed by the model (which was then compared to six year olds). Model-child comparisons were made across six different nonword repetition studies, which, between them, tested children at seven different ages, on six different sets of nonwords that varied across a range of characteristics such as length, consonantal complexity, and phonotactic probability. In total, there were 38 datapoints for model-child comparisons. CLASSIC proved extremely accurate at modeling children's performance. For example, the model's repetition accuracy for individual nonwords significantly correlated ($r = 0.72$) with those of children; overall nonword repetition performance for the model was within 7% of the children and fit over 85% of the child data (33 of 38 datapoints); nonword repetition performance strongly correlated with vocabulary size (as occurs in child studies); and the model and children showed the same effects of length, consonantal complexity and phonotactic probability.

Jones and Macken (2015) also used the same CLASSIC model, trained on a large sample (half a million utterances) of adult-directed speech and written literature, to examine performance in digit span, the archetypal measure of verbal short-term memory. Their computational and empirical studies examined a long-known and hitherto unexplained oddity in digit span performance, namely that span for random sequences of digits is consistently far greater than span for random sequences of other stimuli such as words. Jones and Macken (2015) showed that such 'digit span superiority' can be explained by the fact that random sequences of digits are far more prevalent within natural language, with CLASSIC showing the same digit superiority effects as adults. CLASSIC performed within 4% of the adult span size data across three different types of stimuli (digits, words, and a mixture of the two). Moreover, the model predicted a critical difference between digit lists and word lists at list length five, because digit lists at this length could be coded using far fewer chunks than for word lists. This is precisely where adults faltered for word lists, with 60% failing at this list length compared to only 26% of the same adults failing for digit lists.

Note that CLASSIC has no knowledge of syntax or semantics; it is learning sequence information from naturalistic language input. As such, it is not a fully-realized model of the complete language learning process (e.g., it is unable to produce novel utterances, though see Freudenthal, Pine, Jones, & Gobet, 2015, for a productive model based on a similar architecture). Nevertheless, previous work with CLASSIC shows that it is capturing critical aspects of linguistic knowledge within its learning, and that it is doing so for both sublexical and lexical/multi-word sequences.

3. Study 1: Modeling the effect of input quantity and lexical diversity on Language-related learning using CLASSIC

In Study 1, we investigated the information learned within the model, and how this changed across development, when the input was maternal utterances, when the input was artificially manipulated to increase lexical diversity (keeping quantity constant) and when the input was artificially manipulated to increase quantity (keeping lexical diversity constant).

3.1. Method

3.1.1. Input to the model

The Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001) on CHILDES (MacWhinney, 2000) contains transcripts for 12 children, each comprising 34 h of mother-child interactions recorded over a one-year period. All children began the study aged between 22 and 24 months and were recorded for two hours every three weeks for a year. To increase sample

size to N = 16, four additional children who were of the same age as the children in the Manchester corpus were added: Eleanor and Fraser from the MPA-EVA corpus, Thomas from the Thomas corpus (for both, see Lieven, Salomo, & Tomasello, 2009) and Lara from the Lara corpus (Rowland & Fletcher, 2006; all corpora available on CHILDES, see MacWhinney, 2000). Significantly more data were available for these corpora than for the Manchester corpus children, so we used random sampling to extract thirty-four hours of transcripts for these children. The mean number of utterances across the 16 mothers was 24,385 (SD = 6231).

All maternal utterances addressed to the 16 children were coded phonemically by converting each constituent word to its phonetic equivalent. Note that the phonemic input to the model involved automatic conversion of the utterance string into its phonemic representation using a word-phoneme database. This means the number of utterances for each mother are slightly lower than in the original transcripts. The maternal input is hereafter labelled 'Mother'. Table 2 shows the characteristics of the input.

Artificial manipulations were then made to the Mother inputs in order to either increase the lexical diversity (increasing word types) or increase the input quantity (increasing word tokens) while keeping other properties of the input identical across the input sets. For the Diversity manipulation, we considered replacing half of each mother's utterances with those of another mother from our corpus set (since all are parts of ongoing conversations aimed at 2–3-year-old children). However, this did not significantly increase the number of word types. The Diversity manipulation was therefore created by substituting half of each mother's utterances (taken at random) with utterances addressed to 3–5 year old children, which are more diverse in terms of word types. Utterances were taken from the Forrester (Forrester, 2002), Lara (Rowland & Fletcher, 2006), Smith (Smith, 1973) and Thomas (Lieven et al., 2009) corpora from the English UK transcripts on CHILDES. The sample taken was pseudo-random to ensure that we matched the number of word tokens from the original substituted mother utterances as closely as possible (i.e., MLU and number of word tokens were identical to the original mother utterances), with each mother having a different pseudo-random set of input utterances. In sum, the Diversity manipulation matched the Mother input for input quantity (number of word tokens) but differed in lexical diversity (number of word types).

The Quantity manipulation was created by pseudo-randomly duplicating a subset of utterances from each of the original mother inputs (thereby keeping number of word types constant), again distributing the utterances across the mothers' inputs. The proportional increase in word tokens matched as much as possible the proportional increase in word types in the Diversity input. Although number of utterances is not one of our measures, we also kept the proportional increase in utterances consistent with the increase in word tokens, because the model is presented with the input one utterance at a time (mean utterances = 40,537, SD = 12,777). The Quantity manipulation for each mother therefore matched the Mother input for lexical diversity (number of word types) but differed in input quantity (number of word tokens and utterances).

Each input was presented to the model one utterance at a time and processed in line with the model description above. The model's learning was tested by measuring the information contained within its chunks (sublexical sequences, sequences corresponding to lexical items, multi-word sequences). We measured learning at 5% intervals (i.e., when the model had seen 5% of the input, 10% of the input etc.) to assess learning over developmental time.

The probabilistic processing constraint within the model means that, while on average 4.5 chunks are accessed, this will vary from utterance to utterance, which means results will differ slightly each time the model is run. We therefore performed ten independent model runs for each input type (Mother, Diversity manipulation, Quantity manipulation) for each

Table 2
Input characteristics. All figures are in 1000s except MLU. The Quantity manipulation increases the number of word tokens by duplicating maternal utterances, thus having the same number of word types as the Mother. The Diversity manipulation increases the number of word types by replacing maternal utterances with those from input aimed at older children, thus having the same number of word tokens as the Mother. Both the duplicated maternal utterances and the utterances aimed at older children were pseudo-randomly selected such that MLU remained the same as per the maternal input. The italicized rightmost two columns indicate the changed metrics for the artificially manipulated inputs.

Child	Mother/Diversity: Tokens	Mother/Quantity: Types	All inputs: MLU	*Diversity: Types*	*Quantity: Tokens*
Anne	126.3	2.8	3.8	*5.3*	*234.4*
Aran	169.5	4.2	5.0	*6.3*	*255.5*
Becky	89.1	2.4	3.6	*4.4*	*162.6*
Carl	79.0	2.1	3.7	*4.1*	*150.6*
Dominic	120.3	2.6	3.7	*5.0*	*234.0*
Gail	94.8	3.7	3.8	*5.1*	*131.4*
Joel	95.1	3.6	3.8	*5.0*	*132.5*
John	72.3	2.7	4.1	*4.2*	*112.5*
Liz	69.8	2.6	4.0	*4.1*	*112.7*
Nicole	105.2	3.1	3.9	*4.9*	*165.0*
Ruth	126.5	2.9	3.8	*5.2*	*231.4*
Warren	71.7	3.0	4.4	*4.4*	*106.6*
Eleanor	52.5	2.1	3.4	*3.5*	*88.8*
Fraser	77.8	2.5	3.6	*4.1*	*129.2*
Lara	73.4	2.5	3.5	*4.1*	*120.9*
Thomas	126.3	3.6	5.1	*5.5*	*191.3*
M	*96.9*	*2.9*	*4.0*	*4.7*	*160.0*

of the mothers (N = 16) and averaged performance across the ten runs. For example, the same mother input for Anne was presented to the model ten times and all mother performance measures for Anne were averaged across these ten independent runs.

3.2. Results

We first analyzed the effect of the different input types on the model's learning at the end of the learning process. Table 3 shows the mean number (SD) of chunks learned by the model at the end of the learning period for each type of input. For all items, it is worth noting that despite the probability of learning a chunk being 1.0, the model actually learns far fewer than one chunk per utterance regardless of input type (e.g., only 71% of the Diversity input utterances result in the learning of a new chunk). In part this is because the processing constraint reduces learning opportunities but it is also because single-word utterances, which are learnt quickly, account for 26% of utterances across all inputs (once the word is learned, no further learning can take place on subsequent repetitions).

We analyzed the data using four one-way ANOVAs, with input type (Mother/Diversity/Quantity) as the independent variable; the DVs were, respectively, all chunks, chunks corresponding to a sublexical sequence, chunks corresponding to a lexical item, and chunks corresponding to a multi-word sequence. For each input type, N = 16 because there are 16 different children and therefore 16 different inputs.

There was an effect of input type for all chunks ($F(2,45) = 3.59$, $p = 0.036$, $\eta_p^2 = 0.14$), indicating that the model learned more chunks overall from the Quantity input than the Mother input ($p = 0.015$) and the Diversity input ($p = 0.048$) with no significant difference between the Diversity and Mother inputs ($p = 0.620$). However, this result masks very different patterns across the different types of information held in a chunk (sublexical sequence, lexical item (word), multi-word sequence). For lexical items (words), the significant effect of input type ($F(2,45) = 11.81$, $p < 0.001$, $\eta_p^2 = 0.34$) showed that the model learned significantly more chunks that corresponded to lexical items from the Diversity input than from both the Quantity ($p = 0.005$) and Mother input ($p < 0.001$), with no significant difference between the Quantity input and the Mother input ($p = 0.072$). A similar pattern was seen for sublexical sequences ($F(2,45) = 18.60$, $p < 0.001$, $\eta_p^2 = 0.45$), with significantly more chunks containing sublexical sequences learned from the Diversity input than from both Quantity and Mother input (both $ps < 0.001$), with no significant difference between the latter ($p = 0.618$).

The model exposed to the Quantity input learned more chunks corresponding to multi-word sequences ($F(2,45) = 4.79$, $p = 0.013$, $\eta_p^2 = 0.18$) than both the Diversity and Mother inputs ($p = 0.009$ and $p = 0.012$ respectively), with no significant difference between the latter two ($p = 0.902$). Overall, across all three input types, the model learned more multi-word sequences than lexical items or sublexical sequences. This is not too surprising when one considers the maternal input averages 96,900 word tokens but only 2900 types (i.e., a type:token ratio of 1:33). That is, the model has many opportunities to learn vocabulary items, and once it has done so, subsequent learning will involve multi-word sequences.

A different pattern emerges when we consider how the model learned over time (development). Figs. 2–4 show how the different types of information contained in a chunk changed over time by charting learning over the course of the input. For chunks corresponding to lexical items (Fig. 2) and sublexical sequences (Fig. 3), the Quantity input outperformed the other inputs early in learning (i.e., when only a small amount of the input had been processed) but was quickly superseded by the Diversity input. For chunks corresponding to multi-words (Fig. 4), however, the Quantity input consistently outperformed both the Diversity and Mother inputs throughout development.

For chunks corresponding to lexical items and sublexical chunks, we analyzed the strength of the crossover effect by comparing the number of chunks learned from the different inputs at appropriate points before and after the crossover in a two-way ANOVA. We first determined where the crossover occurred between Diversity and Quantity inputs; we then noted the point at which the Quantity input was maximally superior to Diversity before the crossover ('early in learning'); and, finally, we sampled the point after the crossover that was the same distance away in time. In other words, we sampled from points equidistant before and after the crossover, with the distance determined by the point of maximal difference between inputs before the crossover. For example, the data for lexical items shows that Quantity was maximally superior to Diversity after 5% of the input had been seen, with the crossover occurring after 40% of the input had been seen. Thus, we sampled from the Diversity and Quantity inputs after 5% of the input was seen and after 75% of the input was seen. A 2 (input type: Diversity or

Table 3
Mean number of chunks (SDs) learned by the end of the learning process for each different type of input. A chunk corresponds to a phoneme or sequence of phonemes that has been learned from the input, and as such can contain sublexical sequences (phoneme or phoneme sequences that do not correspond to a word or sequence of words), a lexical item (phoneme or phoneme sequence that corresponds to a word), or a multi-word sequence (phoneme sequence that corresponds to more than one word).

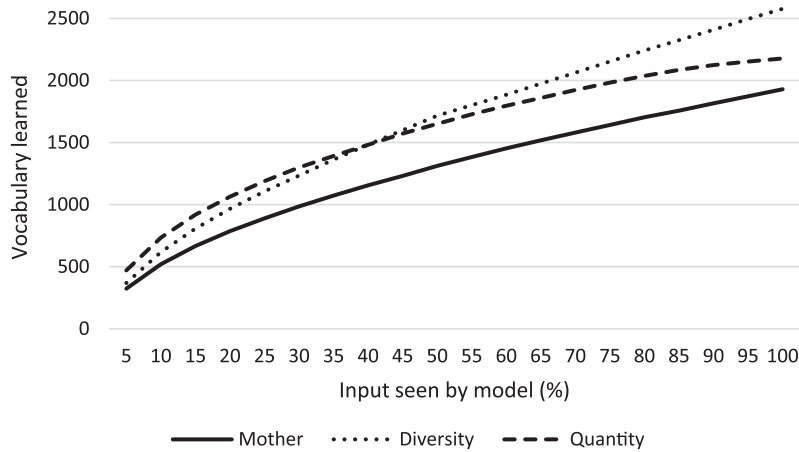| Chunk contents	Input type		
	Mother	Diversity	Quantity
All	16,680 (4148)	17,480 (4278)	20,729 (5071)
Lexical items	1929 (362)	2577 (403)	2177 (376)
Sublexical sequences	1198 (167)	1527 (163)	1228 (176)
Multi-word sequences	13,553 (3756)	13,375 (3740)	17,323 (4661)

Fig. 2. Number of chunks learned that corresponded to lexical items (vocabulary) for each type of input at different stages of the model's learning (N = 16 mothers for each input type). Relative to the maternal input, the diversity input increases word types while keeping word tokens constant; the quantity input increases word tokens while keeping word types constant.
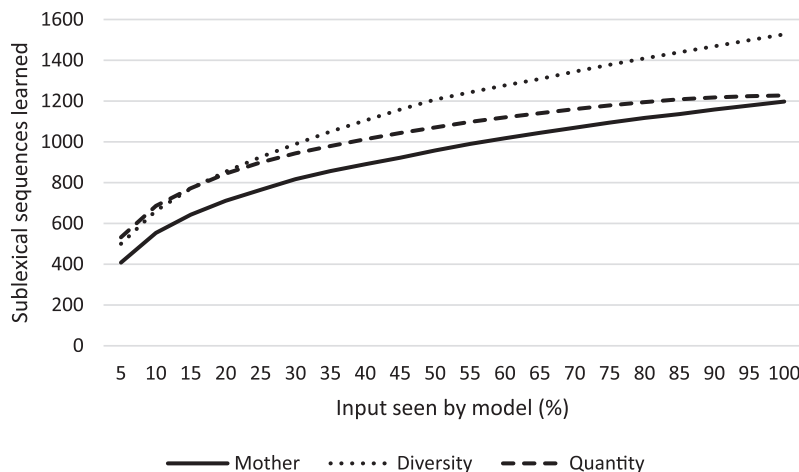

Fig. 3. Number of chunks learned that corresponded to a sublexical sequence for each type of input at different stages of the model's learning (N = 16 mothers for each input type).

Quantity) × 2 (location: pre-crossover, post-crossover) ANOVA showed that the interaction was significant for both lexical sequences ($F(1, 30) = 8.19$, $p = 0.008$, $\eta_p^2 = 0.21$) and sublexical sequences ($F(1, 30) = 18.22$, $p < 0.001$, $\eta_p^2 = 0.38$). Early in learning, the model learned more chunks corresponding to lexical items and sublexical chunks from the Quantity input than the Diversity input. Later in learning, this pattern was reversed.

In sum, for lexical and sublexical knowledge, input quantity leads to faster learning early in the learning process (i.e., when a small amount of the input has been processed), with lexical diversity leading to faster learning subsequently. Input quantity is advantageous throughout for multi-word knowledge.

3.2.1. Parameter manipulation

To determine the effect of the parameter settings of the model (processing limit averaging 4.5 items, the recency assumption, chunk learning probability of 1.0), the model was re-run at a range of different parameter settings (processing limitation 2.25, 3.00, 3.75; probability of learning 0.25, 0.50, 0.75) and also without the recency assumption. The same pattern was found for lexical items and sublexical items, so we present only the data for lexical items here. For simplicity, we do not present data from the Mother input type. Fig. 5 shows the effects of the different processing limitations, Fig. 6 shows the effect of removing the recency bias, and Fig. 7 shows the effects of varying the probability of learning a new chunk. It is quite clear across the figures that most reductions in parameter settings simply shift the crossover point later in development. Lower parameter settings mean that the models require greater quantities of input before an input rich in diversity can gain the upper hand. Removing the recency bias has little effect on the pattern of results: slightly fewer words are learned overall
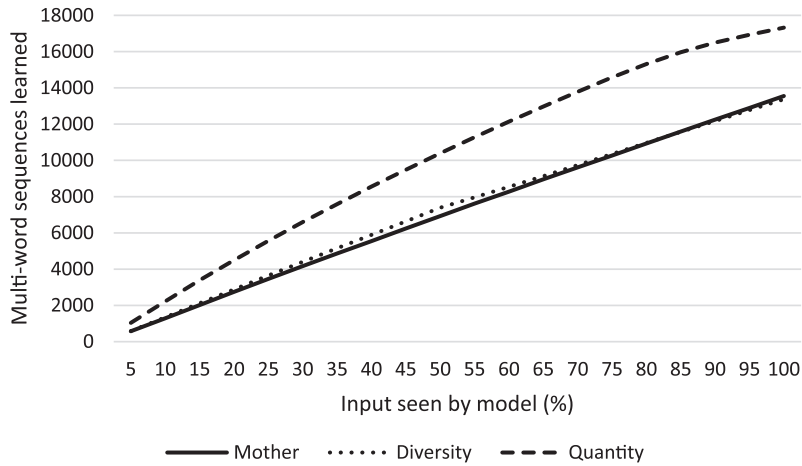
Fig. 4. Number of chunks learned that corresponded to a multi-word sequence for each type of input at different stages of the model's learning (N = 16 mothers for each input type).
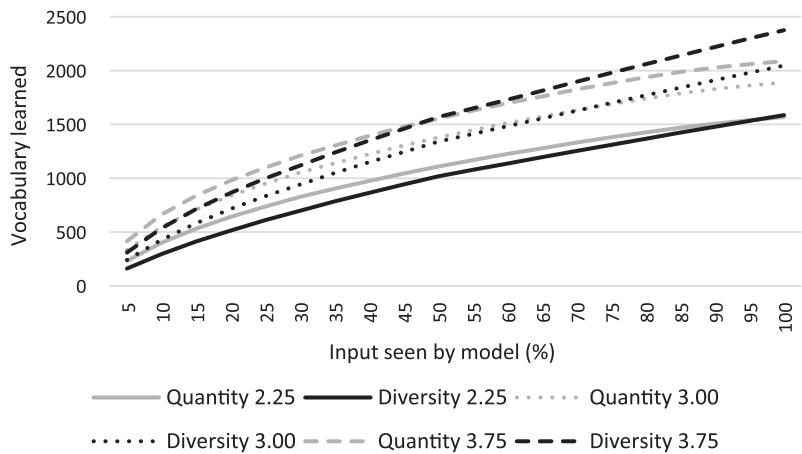

Fig. 5. Effect of Diversity and Quantity inputs at different processing limitations (ranging from 2.25 to 3.75; the original setting was 4.50), for different stages of the model's learning.
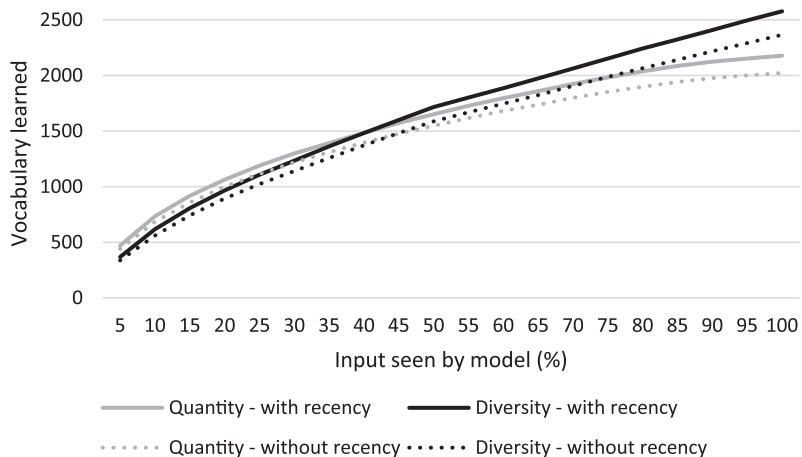

Fig. 6. Effect of Diversity and Quantity inputs with and without the recency bias in the model, for different stages of the model's learning.
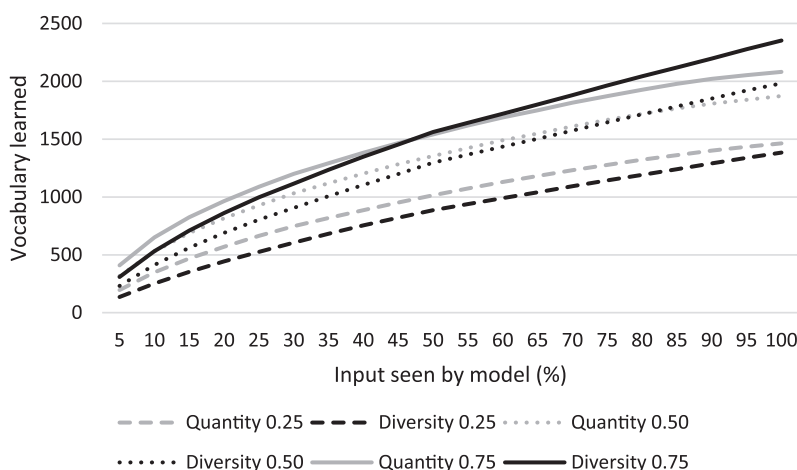
Fig. 7. Effect of Diversity and Quantity inputs for different probabilities of learning a new chunk (probabilities ranging from 0.25 to 0.75, the original setting was 1.00), for different stages of the model's learning.

but the point of crossover is similar to that for the original recency-biased model and Diversity gains an advantage in both models before half of the input has been seen.

Only when the probability of learning a new chunk was very low (0.25) did the Diversity input never surpass the Quantity input by the end of the learning run (see Fig. 7). At this probability, because of the processing constraint, two adjacent chunks in a sequence must be encountered and accessed an average of four times in order for the sequence to be learned as a new chunk. As a result, chunk knowledge in these models was low: the Diversity and Quantity models learned an average of 5851 and 8193 chunks respectively (learning one chunk every 4–5 utterances). In essence, at low probabilities for learning, the model needs greater quantity of input. Thus, larger input sets are required to examine whether very low probabilities for learning will yield the same pattern of effects as we see for other settings. For multi-words, irrespective of parameter settings, the Quantity input always resulted in the model learning more multi-word chunks than the other inputs, for all time points.

3.3. Discussion

Early in learning, the model consistently showed an initial advantage for the Quantity input. However, for lexical and sublexical learning, though not multi-word word learning, the Diversity input quickly superseded the Quantity input, providing a superior learning environment by the end of the learning process. These developmental changes occurred despite the fact that the model has no developmental parameters; while there is a processing constraint, this is fixed to always average 4.5 chunks. Performance differences over time therefore arose from changes in chunked knowledge accrued over time based on the input presented to the model.

The fact that the model learns more *lexical items* from the Diversity than the Quantity input is interesting. Although the Diversity input contains more word types than the Quantity input, this is not the reason for its superiority. In fact, neither model is at ceiling in its performance (75% of the Quantity input lexical items are learned compared to 55% of the Diversity input), and the Quantity input offers more opportunity to learn lexical items (because items are repeated more often). Instead, the Diversity input yields superior learning in the model because vocabulary items are comprised of a limited number of phonemes (~44 for standard British English) and phoneme combinations. The Diversity model learns a large number of these sublexical sequences, and thus has a larger repertoire of chunked sublexical knowledge that can be used to code the input utterances (and the lexical items therein). This increases the likelihood of learning taking place on those items. For example, the model that has already learned the sublexical sequences involved in *make* (m/eɪ/k) will more easily learn related words such as *lake* (l/eɪ/k) because of the shared sublexical chunks across items (similar effects are also seen in children, Coady & Aslin, 2003). In other words, by having a greater amount of sublexical knowledge at its disposal, the Diversity model is more able to process and learn vocabulary items. This is an issue we will return to in Study 4 when we examine how the model learns new words. For the moment, since the conclusions thus far center on a modeling environment, we now examine whether these effects are also seen in children.

4. Study 2: Effects of input quantity and diversity in children

The model above predicts that the quantity of the children's input will initially be the strongest predictor of their vocabulary growth, but that the lexical diversity of their input will eventually overtake input quantity as the strongest predictor. In

Study 2, we investigated this prediction by testing whether individual differences in the quantity and diversity of children's input predicted individual differences in vocabulary growth.

4.1. Method

4.1.1. Participants

The participants were the same 16 children whose input data were used in the model above. As in the model, the input data consisted of maternal utterances only. We now also include the children's utterances (child data) in order to assess child vocabulary size.

To assess the effect of the input on early and late learning separately, we divided the sample in half, into Time 1 and Time 2 data. In order to avoid discourse effects inflating the correlations between input and child data, we extracted child and adult data from different transcripts. At time 1, the input data were extracted from the first 15 h of transcripts (average child age at transcript 1 = 22 months, range = 21–24 months) and the child data from two subsequent hours of transcripts (transcripts 16 and 17; average age = 28 months, range = 25–30 months). At time 2, input data was extracted from transcripts 18–32 (average age at transcript 18 = 28 months, range = 26–30 months), and the child data was extracted from transcripts 33 and 34 (average age = 33 months, range = 32–36 months).

4.1.2. Coding

For the input data (maternal utterances) we used the same measures as for the model: number of word tokens (quantity) and number of word types (lexical diversity). However, unlike for the model, we do not have a direct measure of the children's vocabulary size, so we approximated vocabulary size using the VOCD (VOCabulary Diversity) measure from the CLAN programs (MacWhinney, 2000). VOCD measures the diversity of a participant's vocabulary in a sample of naturalistic speech, which acts as a proxy for vocabulary size, providing a way to compare the relative size of children's vocabulary from naturalistic data. Unlike other measures of vocabulary diversity (e.g., type-token ratio), VOCD controls for sample size by taking into account the fact that samples with larger numbers of tokens yield lower values for type-token ratio and vice versa. The program analyzes the probability of new vocabulary being introduced with an increase in sample size and bases its VOCD score on the type-token ratio versus token curve calculated from data for the transcript as a whole (for a description of VOCD, see Malvern, Richards, Chipere, & Durán, 2004). All measures, for both mothers and children, were extracted from the data using the CLAN programs in CHILDES.

4.2. Results and discussion

The descriptive statistics are shown in Table 4. Statistical analyses were performed in R (R Core Team, 2012) using log transformed data, though raw data are presented in descriptive statistics. Table 5 shows the correlations between all measures. For the input, there were strong correlations between individual measures across time and medium to strong correlations across different measures. There were medium correlations between child time 1 and 2 VOCD scores ($r = 0.47$, N = 16, $p < 0.05$). This suggests good continuity over time for the input quantity and diversity measures, and the child's language learning measures.

There was no significant correlation between input quantity and child vocabulary at either time. In fact, there were negative correlations between quantity and Time1 VOCD scores. Note that the children's Time1 VOCD data were skewed, even when log transformed, by an outlier (Ruth's Time1 VOCD score was only 18.83, compared to a mean across the other 15 children of 64.90 (SD = 11.23)). Nevertheless, even without Ruth (see Table 5), the negative correlations between quantity and Time 1 VOCD scores remain.

As predicted, there were significant correlations of moderate strength between input diversity measures (Diversity Time1, Diversity Time2 and Overall Diversity) and the children's time 2 VOCD scores (bolded in Table 5). This contrasts to the correlations between the input diversity measures and time 1 VOCD scores, which were all small and non-significant.

We ran three regression models to investigate the significant effects of input diversity on Time2 VOCD scores in more detail (see Table 6). Model 1 included Overall Diversity (Time 1 + 2) and Time1 VOCD as predictors, and Time2 VOCD as the dependent variable. Both predictors together explained 40% of the variance in Time2 VOCD scores, and both were inde-

Table 4

Means and standard deviations (in parentheses) across all 16 mother-child dyads. Maternal quantity and diversity at Time 1 is calculated from transcripts recorded when the children were between 22 and 28 months (average age), and at Time 2 from transcripts recorded between 28 and 33 months (average age). To avoid discourse effects, child vocabulary (approximated by VOCD) was calculated from separate transcripts at average age 28 months (time 1) and 33 months (time 2).

Measures		Time1	Time2	All (T1 + T2)
Input	Mean number of word tokens (Quantity)	48,796 (12,972)	49,024 (15,419)	102,641 (20,813)
	Mean number of word types (Diversity)	2191 (416)	2402 (544)	3209 (676)
Child	VOCD (vocabulary size)	62.02 (15.82)	73.96 (12.71)	
	Mean MLU	2.56 (0.41)	3.13 (0.32)	

Table 5
Pearson's (R) correlations between log transformed variables, primarily to examine whether maternal input quantity at time 1 predicts child vocabulary at time 1, and whether maternal lexical diversity at time 2 predicts child vocabulary at time 2. All correlations are one-tailed so only positive (predicted) significant relationships are flagged. Predicted significant correlations are in bold.[1]

	Time	Quantity: No. word tokens			Diversity: No. input word types			Child Vocab. (VOCD)		
		1	2	All	1	2	All	1	1 (no Ruth)	2
Quantity: Number of word tokens	1	1.00	0.88**	0.98**	0.67**	0.73**	0.66**	−0.26	−0.26	0.26
	2		1.00	0.96**	0.59*	0.80**	0.69**	−0.34	−0.22	0.18
	All			1.00	0.62*	0.77**	0.67**	−0.32	−0.26	0.20
Diversity: Number of word types	1				1.00	0.88**	0.95*	0.10	−0.06	**0.47***
	2					1.00	0.97**	−0.03	0.03	**0.43***
	All						1.00	0.03	0.02	**0.48***
Child Vocabulary (VOCD)	1							1.00		0.43
	2									1.00

[1] * $p < 0.05$, ** $p < 0.01$.

Table 6
Results of multiple regressions examining the extent to which diversity and child vocabulary at time 1 predict child vocabulary at time 2 (child vocabulary is approximated by VOCD).

Model	R2	AdjR squared	Residual SE	beta	P[1]
Model 1	0.40	0.31	0.14	–	0.02*
Overall Diversity				0.39	0.02*
VOCD Time 1				0.21	0.04*
Model 2	0.37	0.28	0.15	–	0.02*
Diversity Time1				0.42	0.04*
VOCD Time1				0.19	0.05
Model 3	0.39	0.30	0.15	–	0.02*
Diversity Time 2				0.36	0.03*
VOCD Time 1				0.22	0.03*

* Significant at p < 0.05, one tailed.

pendently significant predictors. Models 2 and 3 that, respectively, included Time1 and Time2 Diversity scores in place of Overall Diversity yielded similar results.

Finally, we ran simple paired correlations to determine whether there was a relationship between the vocabulary of the children at Time2 (VOCD) and the vocabulary of the 16 different mother-input models (each trained on the input of one child's mother) at the end of the learning period. Without Ruth (as an outlier), child vocabulary and model vocabulary showed a medium-strength correlation, but this was non-significant at p < 0.05 ($r = 0.43$, N = 15, $p = 0.055$, one-tailed). This analysis would have to be replicated, with increased power, before strong conclusions could be drawn.

In sum, as predicted by the model, the children's vocabulary at the end of the learning period (time 2) was better predicted by the lexical diversity of the input than by the quantity of input to which children were exposed. However, unlike in the model, there was no effect of quantity at the earlier time period (time 1) and this is an issue we return to in the general discussion. We now return to the model to determine the effects of lexical diversity and input quantity on performance in language-related tasks.

5. Study 3: Testing the model's performance on two tasks that are highly predictive of language ability

In Study 1 we showed that the model learned more chunks that corresponded to sublexical sequences and lexical items (words) from the Diversity input than the Quantity input. In Study 2, we showed that the latter effect held across individual English-learning children; children who heard a more lexically diverse input were able to produce a larger number of different words. In Study 3, we tested whether models that have learnt from an input rich in lexical diversity also perform better on more formal tests that are highly predictive of language ability: (1) a nonword repetition test and (2) a sentence recall test, both of which predict both typical and atypical language performance (e.g., Alloway, Gathercole, Willis, & Adams, 2004; Archibald & Joanisse, 2009; Hoff et al., 2008; Thal, Miller, Carlson, & Vega, 2005). Testing the model on these tasks represents a good method to examine which of input quantity and lexical diversity are likely to result in optimal language learning.

5.1. Method

5.1.1. Nonword repetition
We used the Children's Test of Nonword Repetition (CNRep, Gathercole & Baddeley, 1989). When administered to children, this test involves speaking aloud nonwords, one at a time, with the child's task being to repeat the nonword (scoring

1 for a correct repetition and 0 for an incorrect one). The test comprises 20 nonwords containing consonant clusters (e.g., **sl**adding) and 20 without consonant clusters (e.g., rubid). Each set of 20 nonwords contains five nonwords at each of two, three, four and five syllables in length. We excluded the five syllable items because they are not appropriate for the young children to whom the input is directed. Since the CNRep is generally considered to contain a number of nonwords that are 'wordlike' (e.g., Graf Estes, Evans, & Else-Quest, 2007), we also included the 'non-wordlike' two, three and four syllable nonwords from Jones et al. (2010). These vary in phonotactic probability, with six nonwords at each length containing sublexical sequences that occur relatively frequently in English and six at each length that do not. These nonwords also provide a validity check for the model: because nonwords with low phonotactic probability are less supported by existing linguistic knowledge, the model should perform worse for these nonwords than for the CNRep nonwords. In total, the nonword repetition tests use 66 nonwords, 30 from the CNRep and 36 from Jones et al. (2010).

Each nonword was converted to phonemes and presented to the model one at a time. A nonword was considered to be accurately repeated only when all constituent phonemes within the nonword were represented accurately, given the fixed limit on the number of chunks that can be processed at any one time. The nonword tests were presented to the model during learning at 5% intervals and at the end of the learning period.

5.1.2. Sentence recall

Each sentence in the sentence recall subtest of the Clinical Evaluation of Language Fundamentals – Preschool 2 UK (CELF, Wiig, Secord, & Semel, 2006) was converted to a word-delimited phonemic representation and presented to the model in full. The subtest comprises 12 sentences that gradually increase in length from four words upwards. Subtest scoring procedures took place at the word level. A score of 3 was given for each sentence that was accurately represented within the processing constraint of the model; a score of 2 was given when one word was incorrect; a score of 1 was given when 2–3 words were incorrect; and a score of 0 for 4 or more errors. In line with the administration procedures of this test, no further sentences were presented after three consecutive zero scores. The sentences were presented to the model during learning at 5% intervals and at the end of the learning period.

We also administered an additional sentence recall test because the CELF subtest includes only 12 sentences and we wanted to ensure that any effects seen were not the result of a fortuitous input that happened to learn the words and phrases that were included in the CELF subtest. We therefore chose to capitalize on the properties of the inputs that are used for the models. For each input type for each mother (e.g., Anne), we sourced a random sample of 100 sentences from the remaining 15 mothers that did not appear in any of that mother's speech, together with a random sample of 100 sentences from the older input that did not appear in any input (N = 200 novel sentences, mean MLU = 5.84, SD = 2.81). We did this to ensure a fair comparison across the Quantity and Diversity models, because (for example) one may expect a model trained partly on utterances aimed at older children to perform well on sentences aimed at older children. As per the nonword repetition test, a sentence was considered to be accurately recalled when all constituent phonemes within the sentence were represented accurately within the fixed processing constraint. All 200 sentences were presented to the model during learning at 5% intervals and at the end of the learning period.

5.2. Results and discussion

Means and standard deviations for performance on each test once all of the input had been processed are given in Table 7. Fig. 8 shows how nonword repetition performance varies across the two nonword tests. For both wordlike and non-wordlike nonwords, although there was an initial benefit of Quantity, the Diversity input exerted superiority even before half of the input had been seen. This occurred more quickly for the non-wordlike nonwords because these gained greater benefit from the variation in sublexical sequences that the Diversity input affords. However, at the end of learning, there were effects of input type on both nonword tests (wordlike nonwords: $F(2,45) = 18.61$, $p < 0.001$, $\eta_p^2 = 0.45$; non-wordlike nonwords $F(2,45) = 15.68$, $p < 0.001$, $\eta_p^2 = 0.41$). In both cases, the Diversity input outperformed both the Quantity and the Mother input ($p < 0.001$ in all cases) with no significant difference between the Quantity and Mother inputs ($p > 0.364$ or greater).

Table 7
Means and standard deviations (in parentheses) for performance on the nonword repetition tests (% repeated correctly, N = 66), the CELF sentence recall test (minimum score = 0, maximum score = 36), and the recall of novel sentences (% recalled accurately, N = 200) at the end of the model's learning for each different type of input. Relative to the maternal input, the diversity input increases word types while keeping word tokens constant; the quantity input increases word tokens while keeping word types constant.

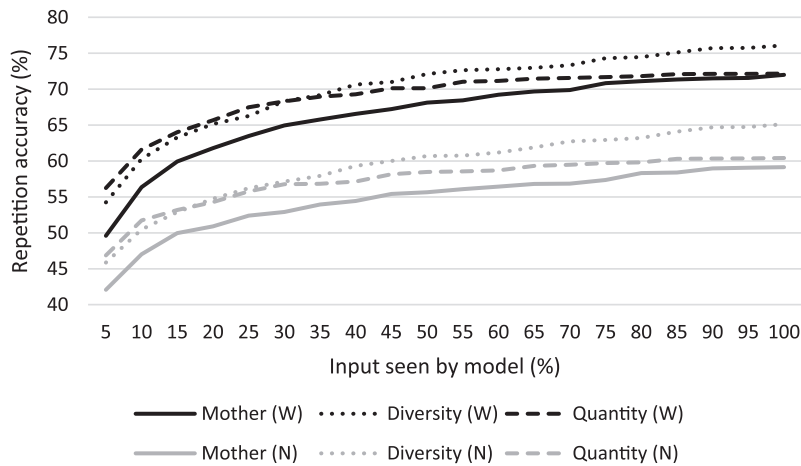| Language test	Input type		
	Mother	Diversity	Quantity
Wordlike nonwords	71.85 (2.19)	76.10 (2.24)	72.13 (2.19)
Non-wordlike nonwords	59.28 (2.43)	65.13 (2.81)	60.31 (4.01)
CELF sentence recall	13.67 (1.79)	16.07 (1.07)	14.50 (1.60)
Novel sentence recall	51.98 (4.22)	56.33 (4.02)	53.97 (4.36)

Fig. 8. Repetition accuracy for wordlike nonwords (W) and non-wordlike nonwords (N) for the different input sets and across different amounts of input seen by the model.

The superior performance of the Diversity model is because of the information in the chunks themselves. Nonword repetition in the model primarily involves the use of sublexical knowledge – because a nonword is novel it will not have already been chunked into a lexical chunk. The Diversity model learns a greater number of sublexical chunks than the Quantity model and is therefore more likely to have already encountered many of the sublexical sequences within nonwords. The Quantity model, on the other hand, is less likely to have encountered some of the sublexical sequences in nonwords and would have to (inefficiently) represent those unfamiliar sequences as a series of individual phoneme chunks. Indeed, at the end of learning, the average size (i.e., number of phonemes) of the chunks used by the Diversity model is 2.28 for wordlike nonwords and 2.16 for those that are not wordlike, compared to 2.10 and 2.02 for the Quantity model. That is, the Diversity model is more able to efficiently represent nonwords because it learns a greater array of sublexical information.

Fig. 9 shows the CELF sentence recall scores across the three input sets. As with the nonwords, Quantity initially outperformed Diversity but before half of the input had been seen, this situation reversed. At the end of learning, there was an effect of input type ($F(2, 45) = 10.21$, $p < 0.001$, $\eta_p^2 = 0.31$) with the Diversity model performing significantly better on the CELF test than the Mother model ($p < 0.001$) and the Quantity model ($p = 0.006$). Fig. 10 shows the sentence recall scores involving novel sentences (N = 200) across the three input sets. The pattern of results is identical to that of the CELF subtest: the Quantity input initially outperformed the Diversity input but before half of the input had been seen, this situation reversed. At the end of learning, there was an effect of input type ($F(2, 45) = 4.71$, $p = 0.014$, $\eta_p^2 = 0.17$) with the Diversity model performing significantly better on sentence recall than the Mother model ($p = 0.004$), and with the results in the predicted direction (albeit non-significant) compared to the Quantity input ($p = 0.064$).

Intuitively, these are perhaps unexpected results – after all, the model will process sentences most efficiently when the sentence can be coded using chunks that correspond to multi-word sequences (since these hold the greatest amount of information about word sequences) and least efficiently when the sentence can only be coded using chunks that correspond to
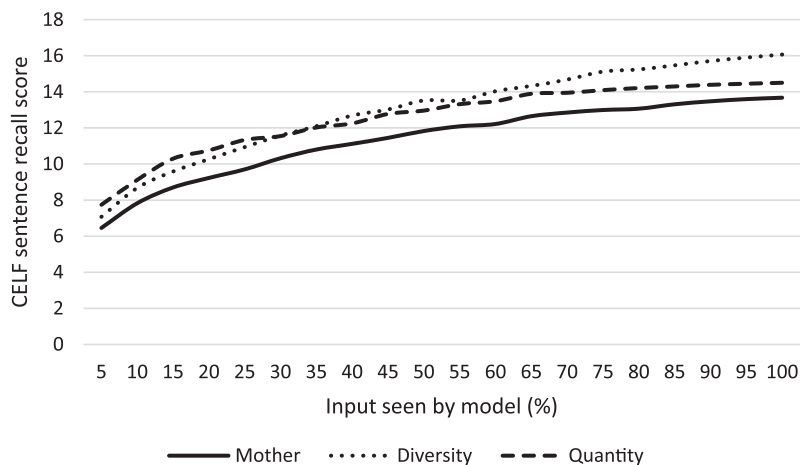

Fig. 9. CELF sentence recall subtest scores for each input set and for different stages of the model's learning.
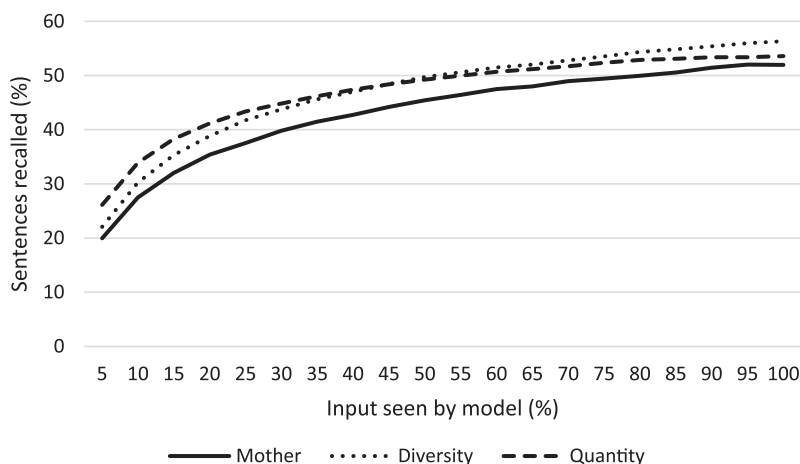
Fig. 10. Proportion of sentences accurately recalled (N = 200) for each input set and for different stages of the model's learning.

sublexical sequences. Since the Quantity model holds the most multi-word chunks, one might expect it to outperform the Diversity model. However, sentence recall involves *novel* sequences of lexical items, and some of those lexical items may not have been encountered before (i.e., bringing to the fore the extent of the model's knowledge of lexical items and sublexical sequences). Take for example the sentence 'James and the giant peach': the model that knows each lexical item as a chunk can code this sentence using at most five chunks, and fewer than five if some of the sequence is known as a multi-word chunk; however, if one lexical item is unknown then not only is the ability to use multi-word chunks more restricted, but also the model would need to fall back on its sublexical knowledge – both of these points would likely significantly increase the number of chunks needed to code the sentence. The greater number of chunks that correspond to lexical items in the Diversity model over the Quantity model therefore allow it to more efficiently process novel sentences because the Diversity model has less need to fall back on sublexical knowledge: for example, when coding the 200 novel sentences, the Diversity model need only use sublexical knowledge for 9% of chunks compared to 14% for the Quantity model. For the Diversity input, therefore, a greater range of sublexical information leads to easier learning of lexical items, and a subsequently greater range of lexical items leads to easier processing of novel sentences.

In Study 3 we have shown that the Diversity model, which learns more sublexical and lexical chunks than the Quantity model, outperforms the Quantity model in representing new verbal material in the form of nonwords and novel sentences. The advantage that the Diversity input affords in relation to processing novel stimuli should also extend to learning new vocabulary, which we tested below in Study 4.

6. Study 4: Effect of input diversity and quantity on learning novel words

In Study 4, we assessed the ability of the models trained on the Quantity and Diversity inputs to learn new words. In the child language literature, there is good evidence for a virtuous circle between the input, child vocabulary size and the speed of new learning. Children who have been exposed to richer input not only possess a larger vocabulary than their peers but also process words more quickly. In turn, this affects the speed with which children learn new words and, ultimately, on the size of their vocabulary in later years. For example, Weisleder and Fernald (2013) showed that 19 month old infants who heard more child-directed speech were quicker to process familiar words presented to them in real time, and had a larger expressive vocabulary by 24 months (see also Hurtado et al., 2008). In fact, the advantages of faster processing speed in young children can still be seen years later; Marchman and Fernald (2008) demonstrated that the speed of spoken word recognition and vocabulary size in a group of 25 month old children predicted the same children's linguistic and cognitive skills six years later, at the age of eight years.

However, it is not yet clear whether it is quantity or diversity that is driving this effect, because it is so difficult to disentangle the two in real speech. The findings from Study 3 above predict that it will be input diversity that is important here, because a more lexically diverse input gives the model a broader repertoire of sublexical and lexical linguistic knowledge, which it can then use to process new words in fewer chunks (i.e. faster) on the first presentation. This means that new words will require fewer presentations before they are fully parsed, and chunked, which means that they will be learned more quickly and efficiently. In Study 4 we test this prediction.

6.1. Method

For each mother (e.g., Anne), we sourced 100 novel words from the remaining 15 mothers that were not seen in the current mother's (e.g., Anne's) input, together with 100 novel words from the older input that did not appear in any of the inputs
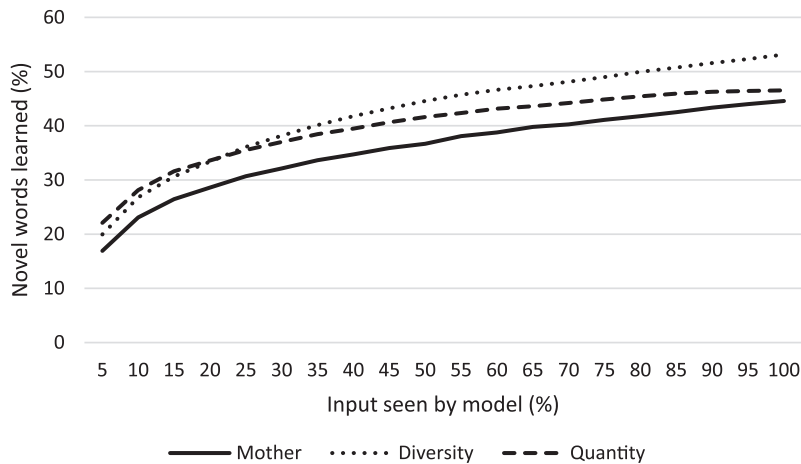
Fig. 11. Novel words learned (%) for each input set and for different stages of the model's learning.

(N = 200, mean phonemic length = 5.74, SD = 1.79). At each stage of the model's learning, we presented each individual novel word to the model and recorded how many chunks were required to code the word. We counted each word as learnt if the model were able to represent the word using two chunks only. We used two chunks as the criteria for learning because of the way the model learns; since the model learns a new chunk for each adjacent set of chunks, if the novel word can be represented using two chunks, the model would learn it on the first presentation if both chunks were accessed. Novel words were presented individually to the model during learning, at 5% intervals, using the same model and inputs as in Study 1. As with the nonword repetition and sentence recall tests, no learning occurred in the model when presenting each new word, because doing so would add new information to the model's existing chunked knowledge, which would impact on the model's performance of the remaining novel words and linguistic input. The dependent measure was the proportion of novel words that the model was able to learn (i.e., represent in two chunks) at each stage of the model's input.

6.2. Results and discussion

Fig. 11 shows the proportion of novel words that the model was able to learn across the learning cycle.

As with the nonword and sentence recall data, the Quantity input had an early advantage but the Diversity input quickly gained superiority before half of the input was seen. By the end of learning, there was a significant effect of input type ($F(2,45) = 14.71$, $p < 0.001$, $\eta_p^2 = 0.40$) with the Diversity model learning a greater proportion of the novel lexical items ($M = 53.16$, $SD = 4.18$), than both the Quantity input ($M = 46.53$, $SD = 4.97$) and Mother input ($M = 44.56$, $SD = 4.95$, both $ps < 0.001$), but with no significant difference between Mother and Quantity inputs ($p = 0.244$). Once again, a larger store of sublexical and lexical chunks enabled the Diversity model to code the novel words more efficiently. In sum, increasing the lexical diversity of an input rather than increasing input quantity ultimately results not only in a larger vocabulary (Study 1) but faster learning of novel words.

7. General discussion

We used a computational model to compare the quality of the language input against its quantity by manipulating the amount of lexical diversity (operationalized as number of different word types) and the amount of input quantity (operationalized as number of word tokens). In Study 1, early in learning, increases in input quantity outperformed increases in lexical diversity for both lexical and sublexical knowledge. However, mid-way through the model's learning, increases in diversity became more important than increases in input quantity, exerting a superiority that increased in strength until all of the input had been processed. In contrast, input quantity consistently outperformed lexical diversity for multi-word knowledge. In Study 2, we then tested these findings against child data, confirming that at time 2, though not at time 1, an input rich in lexical diversity led to more lexical learning than an input rich in quantity. In Study 3, two well-established predictors of language proficiency were then applied to the model (nonword repetition and sentence recall) and, in both cases, increasing the diversity of the input rather than the quantity of it led to significantly better language outcomes. In Study 4, we demonstrated that the Diversity model not only learnt more words than the other models, it was also capable of learning novel words more quickly too. A significantly greater number of novel words were learned when the model experienced a lexically diverse input than when it experienced an input rich in quantity. In sum, the work suggests that while quantity of input is important in early learning, thereafter, it is the diversity of one's language exposure that ultimately drives vocabulary growth.

This study is one of the first direct comparisons of lexical diversity versus input quantity that is not confounded by other variables. We have long known that children who hear richer input learn language faster and learn more vocabulary items overall, but we have not until recently been able to establish what properties of a rich input contribute to this effect. The results here confirm those of Rowe (2012), which suggest that while input quantity is important during the very early years of language learning (up to ~30 months), thereafter it is crucial for the child to hear a range of utterances and vocabulary items. In other words, the definition of optimal input can change over development, depending on the linguistic knowledge that the model (or child) already possesses (see Bohannon & Hirsk-Pasek, 1984, for similar arguments). Eventually, however, both for the model, and for the English-learning children tested here, better outcomes were ultimately reported for more lexically diverse inputs.

However, there are three caveats that must be applied to this conclusion. First, while measures of lexical diversity explain a significant amount of variance in vocabulary learning across children (on average, about 20% in the literature as a whole), the remaining variance also needs to be accounted for. One factor likely to explain additional variance is the child's prior knowledge must play an additional role in later learning. In the model (studies 1, 3 and 4), the amount and type of prior, chunked knowledge affected subsequent performance, and in children (Study 2), VOCD at time 1, combined with lexical diversity, explained 40% of the variance in vocabulary learning. However, many other cognitive factors (e.g. socio-cognitive skills, attention span, working memory), environmental factors (e.g. Weikum, Oberlander, Hensch, & Werker, 2012), and genetic factors (e.g. Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001) are likely to affect the rate of learning too, and will need to be considered in a full model of vocabulary learning.

Second, at very low probabilities for learning (0.25 in the simulations presented), an input rich in lexical diversity did not quite surpass an input rich in input quantity by the end of the model's learning. Quantity is clearly of greater importance when learning is difficult; one first needs a reasonable knowledge base before increases in diversity can be capitalized on. The probability for learning in the original models was set to 1.0 because the available maternal speech samples reflect only a small proportion of the language children hear. Larger speech samples are required to investigate whether the quantity-early/diversity-later pattern holds for low learning probabilities.

Third, we do not yet know how accurately CLASSIC's learning architecture captures real children's vocabulary learning. CLASSIC is a symbolic model that implements a chunk-based learning mechanism that makes graduate changes to the model's knowledge representations on the basis of incoming input. Its mechanism is associative, not error-based, because learning (representational change) is not dependent on the model making erroneous predictions about the upcoming input (unlike, for example, Chang, Dell, & Bock's, 2006, Dual Path model of syntax acquisition). However, the learning in CLASSIC is not associative in the classical sense because the model does not make changes to its representations on every parse of the input (unlike, for example, the word learning model of Fazly, Alishahi, & Stevenson, 2010). CLASSIC learns only when presented with new information, so if an incoming utterance in the model is already represented as one chunk, no learning occurs. In essence, whether learning occurs depends on a combination of the model's state at the time of learning (which is determined by previous learning) and the identity of the incoming utterance (and how similar it is to all previous utterances processed).

CLASSIC's learning mechanism is based on a wealth of evidence about human memory and learning in children and adults (see Gobet et al., 2001) and implements psychologically plausible parameters (e.g. a limitation on the amount of information that children can process at any one time, Gathercole, 2006). However, CLASSIC is clearly not capturing the child data exactly. The correlation between the final output of the model and the children's Time 2 VOCD scores was only moderate and non-significant. In addition, for the children, unlike the model, there was no effect of sheer exposure to language, even at the earlier time period (time 1). This is partly because, as stated above, there are many other factors that influence child vocabulary learning that were not included in the model. It is also, perhaps, because the match between child age and model "age" is not exact; the input we gave to the model constitutes only a fraction of the input that children have heard by the age of two years. However, perhaps more importantly, it is also because the model is not capturing all aspects of the process of learning a word; most notably, it omits the process of learning word meanings as well as the role that other factors such as prosody and syntax may play (see e.g. Christophe, Dupoux, Bertoncini, & Mehler, 1994; Gillette, Gleitman, Gleitman, & Lederer, 1999). Ultimately, to learn a word, one needs to map a particular phonological form (a sequence of phonemes with a particular stress pattern) onto a particular semantic concept, which is itself a complex, multi-dimensional task. The role of the input in this part of the word learning process is not considered in this paper at all and will need to be considered in future work.

In addition, CLASSIC implements just one of many plausible models of the child's vocabulary learning mechanism. In future work it will be important to compare the predictions of our model, in which the input interacts with a chunk-based learning mechanism, with models that implement other types of learning mechanisms. We are not currently aware of any models that have addressed similar questions. However, we are quite optimistic that CLASSIC will make unique predictions about development when compared against other models. For example, it is difficult to see how simple, unmodified one-shot learning models (c.f., Stevens, Gleitman, Trueswell, & Yang, 2016), which store word forms intact (as a unit) on the first hearing, could make similar predictions about the developmental effects we see here (input quantity important early in learning, but lexical diversity ultimately superior), or the effects we see in nonword repetition tests (e.g., nonwords that are phonologically similar to real words are repeated more easily and more accurately by young children, Gathercole, 1995). It is also difficult to see how an error-driven learning mechanism (c.f. Chang et al., 2006; Gasser & Smith, 1998) would yield the developmental effects we see here, even if that mechanism was capable of sequencing or chunking phonemes on the basis of distributional information. In error-based mechanisms, learning occurs only when the model makes a wrong prediction.

Since incorrect predictions are more likely from diverse (unpredictable) input, such models would likely predict an effect of input diversity from the very beginning of learning. Finally, we might predict that models that implement traditional cross-situational associative learning (e.g., Fazly et al., 2010), if adapted to learn the phonological form of words, would learn most quickly from an input that is rich in quantity, because these models require substantial repetition of a word in a variety of contexts before word learning takes place. In future work, it will be important to test our predictions against implementations of different learning mechanisms.

The results from the CLASSIC model also yield three hypotheses about the mechanisms by which children learn words, which make clear predictions that can be tested against child data. First, the results suggest that lexical diversity will be more important than mere repetition (input quantity) in all but the early stages of word learning. Increasing the lexical diversity of the input enabled our model to build a larger store of sublexical chunks that were then available to process new, incoming input. The model learned by coding utterances into as few chunks as possible and combining adjacent chunks on each exposure until the whole word was represented as one chunk. It was therefore a huge advantage for the model to have a large store of sublexical chunks, because it meant that new words could be processed using fewer, larger chunks, which increased the chances of the word being learnt with fewer exposures. In other words, by having a greater repertoire of sublexical knowledge at its disposal, the model trained on the Diversity input could learn new vocabulary items more quickly. The prediction here, therefore, is that language input that is lexically diverse from a relatively early age will support better and faster vocabulary learning, because once a critical level of sublexical information has been learned, children may be ready to exploit that knowledge to great effect. In addition, since varying the learning parameters resulted in a similar pattern of results, but with a later cross-over, the model predicts that children who receive less input should have a longer period in which an input that is rich in quantity, not lexical diversity, leads to faster word learning. We have tested the first of these predictions using the VOCD statistic on children's word production data in Study 2 above, but this measure is only a proxy for vocabulary learning. Further work using more direct measures of vocabulary knowledge (e.g., standardized test scores), testing children who have been exposed to more or less diverse inputs, is needed to test predictions about the influence of lexical diversity and input quantity across development.

Second, the findings suggest a reason why children's performances on nonword repetition and sentence recall tests are such strong predictors of vocabulary learning. Recall that the Quantity input actually led to the model learning a greater total amount of linguistic information than the Diversity input (mostly in the form of multi-word sequences, see Table 3) and yet the model trained on the Diversity input outperformed the Quantity input model on nonword repetition and sentence recall tasks. That is, it is not the sheer amount of learned linguistic information that is important, but what kind of linguistic information is learned. The performance of the model on these tests suggests that a key component of these tasks relates to how the child is able to bring to bear what they have learned from their linguistic experience on the novel stimuli involved. In essence, these tasks are measuring the flexibility of the child in dealing with a novel stimulus and thus both capitalize on variation in learned knowledge. In the case of nonwords this is largely governed by the amount of sublexical information experienced by the rememberer, and in the case of sentence recall this is largely governed by the amount of lexical information experienced by the rememberer. Thus, the prediction here is that children's performance on both of these tests will be determined by their sublexical and lexical knowledge, which in turn will be predicted by the diversity of their input.

Third, the results suggest a reason for the virtuous circle in which children who have been exposed to richer input are able to process words more quickly, which in turn affects the speed with which they learn new words and, ultimately, on the size of their vocabulary in later years (Weisleder & Fernald, 2013). Our results suggest that it is lexical diversity that is responsible for this effect. The model exposed to the Diversity input learned a large number of sublexical and lexical chunks, which enabled it to process incoming new words more quickly (because it could use pre-existing chunked knowledge rather than building representations from scratch, phoneme by phoneme). Thus, as well as predicting that lexical diversity will have stronger correlations with children's speed of processing than quantity of exposure, this paper makes a clear case for why we find a relationship between input richness, processing speed and vocabulary size in children's language learning. Exposure to lexically diverse inputs enable a broad range of linguistic material to be learned, particularly sublexical and lexical material. This allows a greater amount of a novel input to be processed at each exposure, which, in turn, allows more of the input to be learned. Over time, fewer exposures are needed to learn novel words, which leads to faster learning long-term and, ultimately, to a larger vocabulary. If this prediction is upheld when tested with children, it has strong implications for our theories of word learning, but also for interventions for children whose language learning is slower than we would like (e.g. some disadvantaged children, see Fernald, Marchman, & Weisleder, 2013). Interventions that provide children with a lot of repetitive input initially, which is slowly replaced by more lexically diverse input, may lead to more optimal learning; not only exposing children to a greater range of sublexical sequences and to a greater range of words, but also increasing the speed of their subsequent learning, with a subsequent positive effect on language and cognitive development in later childhood.

In summary we have provided a computational analysis of caregiver speech in which maternal utterances were manipulated to increase the amount of lexical diversity (quality) of the input while keeping the amount of language exposure (quantity) constant and vice versa. The analysis showed that an input rich in lexical diversity outperforms an input equivalently rich in quantity for learned sublexical and lexical knowledge, for well-established language tests, and for acquiring words that have never been encountered before.

Acknowledgments

References

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87*, 85–106. http://dx.doi.org/10.1016/j.jecp.2003.10.002.

Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52*, 899–914. http://dx.doi.org/10.1044/1092-4388(2009/08-0099).

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92*, 265–280. http://dx.doi.org/10.1016/j.jml.2016.07.004.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248.

Beals, D. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language, 24*, 673–694. http://dx.doi.org/10.1017/S0305000997003267.

Beckman, M. E., & Edwards, J. (1999). Lexical frequency effects on young children's imitative productions. In M. Broe & J. Pierrehumbert (Eds.). *Papers in laboratory phonology* (Vol. 5, pp. 207–217). Cambridge, UK: Cambridge University Press.

Bohannon, N., & Hirsh-Pasek, K. (1984). Do children say as they're told? A new perspective on motherese. In L. Feagans, C. Garvey, & R. M. Golinkoff (Eds.), *The origins and growth of communication* (pp. 176–195). Norwood, NJ: Ablex.

Bornstein, M. H., Haynes, M. O., & Painter, K. M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language, 25*, 367–393. http://dx.doi.org/10.1017/S0305000998003456.

Bornstein, M. H., & Tamis-LeMonda, C. S. (1995). Language and nonlanguage factors in development of prelinguistic vocalization to linguistic communication. In G. Konopczynski (Ed.), *Is early language performance predictive of later language development?* (pp 81–102). Calais, France: Ortho-Editions.

Cartmill, E., Armstrong, B. F., Gleitman, L. R., Goldin-Meadows, S., Medinac, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences, 110*, 11278–11283. http://dx.doi.org/10.1073/pnas.1309518110.

Catts, H. W., & Kamhi, A. G. (1984). Simplification of /s/+stop consonant clusters: A developmental perspective. *Journal of Speech and Hearing Research, 27*, 556–561. http://dx.doi.org/10.1044/jshr.2704.556.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272. http://dx.doi.org/10.1037/0033-295X.113.2.234.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America, 95*(3), 1570–1580. http://dx.doi.org/10.1121/1.408544.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighborhoods in the developing lexicon. *Journal of Child Language, 30*, 441–469. http://dx.doi.org/10.1017/S0305000903005579.

De Groot, A. D. (1978). *Thought and choice in chess*. Netherlands: Mouton.

Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Psycholinguistic Research, 43*, 149–165. http://dx.doi.org/10.1007/s10936-013-9245-7.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech, 49*, 137–174. http://dx.doi.org/10.1177/00238309060490020201.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition, 2*, 189–220. http://dx.doi.org/10.1207/s15327817la0203_1.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science, 34*, 1017–1063. http://dx.doi.org/10.1111/j.1551-6709.2010.01104.x.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *Macarthur-bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*, 234–248. http://dx.doi.org/10.1111/desc.12019.

Forrester, M. A. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood, 9*, 255–276. http://dx.doi.org/10.1177/0907568202009003043.

Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2015). Simulating the cross-linguistic pattern of Optional Infinitive errors in children's declaratives and Wh- questions. *Cognition, 143*, 61–76. http://dx.doi.org/10.1016/j.cognition.2015.05.027.

Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes, 13*, 269–306. http://dx.doi.org/10.1080/016909698386537.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83–94. http://dx.doi.org/10.3758/BF03210559.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543. http://dx.doi.org/10.1017/S0142716406060383.

Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*, 672–688. http://dx.doi.org/10.1006/jmla.1994.1032.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213. http://dx.doi.org/10.1016/0749-596X(89)90044-2.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*(2), 135–176. http://dx.doi.org/10.1016/S0010-0277(99)00036-0.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236–243. http://dx.doi.org/10.1016/S1364-6613(00)01662-4.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Hearing, and Language Research, 50*, 177–195. http://dx.doi.org/10.1044/1092-4388(2007/015).

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67*, 106–148. http://dx.doi.org/10.1016/j.jml.2012.04.004.

Gupta, P. (2003). Primacy and recency in nonword repetition. *Memory, 13*, 318–324. http://dx.doi.org/10.1080/09658210344000350.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28*, 1096–1105. http://dx.doi.org/10.1037/0012-1649.28.6.1096.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., ... Suma, K. (2015). The contribution of early communication quality to low-income children's language Success. *Psychological Science, 26*, 1071–1083. http://dx.doi.org/10.1177/0956797615581493.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88. http://dx.doi.org/10.1016/j.dr.2005.11.002.

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20-to 24-month-olds. *Journal of Child Language, 35*, 903–916. http://dx.doi.org/10.1017/S0305000908008751.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*, 1–27. http://dx.doi.org/10.1017/S0305000910000759.

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

Hsu, N., Hadley, P. A., & Rispoli, M. (2015). Diversity matters: Parent input predicts toddler verb production. *Journal of Child Language*, 1–24. http://dx.doi.org/10.1017/S0305000915000690.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1217–1232. http://dx.doi.org/10.1037//0278-7393.23.5.1217.

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

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236–248. http://dx.doi.org/10.1037/0012-1649.27.2.236.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*, 343–365. http://dx.doi.org/10.1016/j.cogpsych.2010.08.002.

Jakobson, R. (1941/1968). *Child language, aphasia, and phonological universals*. The Hague: Mouton (1968). Translation by A.R. Keiler of Kindersprache, Aphasie, und allgemeine Lautgesetze. Uppsala: Uppsala Universitets Arsskrift (1941).

Jones, G. (2016). The influence of children's exposure to language from two to six years: The case of nonword repetition. *Cognition, 153*, 79–88. http://dx.doi.org/10.1016/j.cognition.2016.04.017.

Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., & Pine, J. M. (2014). Why computational models are better than verbal theories: The case of nonword repetition. *Developmental Science, 17*, 298–310. http://dx.doi.org/10.1111/desc.12111.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*, 853–873. http://dx.doi.org/10.1111/j.1467-7687.2007.00638.x.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13. http://dx.doi.org/10.1016/j.cognition.2015.07.009.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research, 53*, 1642–1655. http://dx.doi.org/10.1044/1092-4388(2010/09-0222).

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*, 519–523. http://dx.doi.org/10.1038/35097076.

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20*, 481–508. http://dx.doi.org/10.1515/COGL.2009.022.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, F9–F16. http://dx.doi.org/10.1111/j.1467-7687.2008.00671.x.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97. http://dx.doi.org/10.1037/h0043158.

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48*, 439–458. http://dx.doi.org/10.1044/1092-4388(2005/030).

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763–782. http://dx.doi.org/10.1111/j.1467-8624.2005.00876.x.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*, 41–58. http://dx.doi.org/10.1017/S0142716400009863.

R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review, 102*, 305–330. http://dx.doi.org/10.1037/0033-295X.102.2.305.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*, 1762–1774. http://dx.doi.org/10.1111/j.1467-8624.2012.01805.x.

Rowland, C. (2014). *Understanding child language acquisition*. Abingdon: Routledge. doi: 10.4324/9780203776025.

Rowland, C. F., & Fletcher, S. L. (2006). The effect of sampling on estimates of lexical specificity and error rates. *Journal of Child Language, 33*, 859–877. http://dx.doi.org/10.1017/S0305000906007537.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Smolak, L., & Weinraub, M. (1983). Maternal speech: Strategy or response? *Journal of Child Language, 10*, 369–380.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2016). The pursuit of word meanings. *Cognitive Science*, 1–39. http://dx.doi.org/10.1111/cogs.12416.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology, 43*, 454–464. http://dx.doi.org/10.1037/0012-1649.43.2.454.

Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research, 48*, 1481–1495. http://dx.doi.org/10.1044/1092-4388(2005/103).

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*, 127–152. http://dx.doi.org/10.1017/S0305000900004608.

Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences, 109*(Supplement 2), 17221–17227. http://dx.doi.org/10.1073/pnas.1121263109.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143–2152. http://dx.doi.org/10.1177/0956797613488145.

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*, 265–279. http://dx.doi.org/10.1037/0012-1649.37.2.265.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical evaluation of language fundamentals-UK* (2nd ed.). London, UK: Harcourt Assessment.