Trends in Language Acquisition Research 35

CHAPTER 4

Explaining individual differences in children's vocabulary growth

Insights from the Language 0–5 Project

Authors

Caroline F. Rowland | Max Planck Institute for Psycholinguistics | Donders Institute for Brain, Cognition and Behaviour

Gary Jones | Nottingham Trent University

Michelle S. Peter | Great Ormond Street Hospital for Children NHS Foundation Trust

Amy Bidgood | Liverpool John Moores University

Andrew Jessop | University of Liverpool

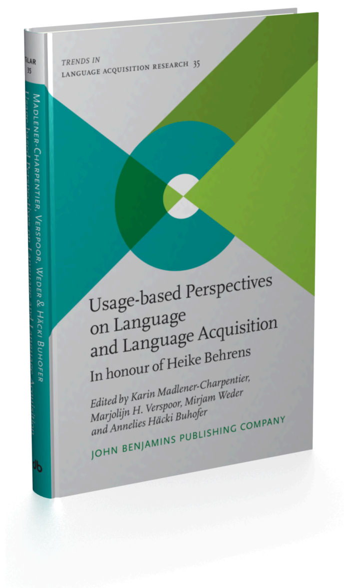
Samantha Durrant | University of Manchester

Paula Stinson | University of Liverpool

Julian M. Pine | University of Liverpool

https://doi.org/10.1075/tilar.35.04row

Available under a CC BY-NC-ND 4.0 license.



Pages 111–140 of
Usage-based Perspectives on Language and Language Acquisition: In honour of Heike Behrens
Edited by Karin Madlener-Charpentier, Marjolijn H. Verspoor, Mirjam Weder and Annelies Häcki Buhofer
[Trends in Language Acquisition Research, 35]
2026. xx, 343 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material, beyond the permissions granted by the Open Access license, written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights

CHAPTER 4

Explaining individual differences in children's vocabulary growth

Insights from the Language 0–5 Project

Caroline F. Rowland,[1,2] Gary Jones,[3] Michelle S. Peter,[4] Amy Bidgood,[5] Andrew Jessop,[6] Samantha Durrant,[7] Paula Stinson[6] & Julian M. Pine[6]

[1] Max Planck Institute for Psycholinguistics | [2] Donders Institute for Brain, Cognition and Behaviour | [3] Nottingham Trent University | [4] Great Ormond Street Hospital for Children NHS Foundation Trust | [5] Liverpool John Moores University | [6] University of Liverpool | [7] University of Manchester

Our goal in this chapter was to test a usage-based computational model of vocabulary learning (CLASSIC). CLASSIC simulates variation in vocabulary growth as a product of interactions between linguistic input quantity and the child's current knowledge state, using an associative learning mechanism with a fixed capacity limited processing window. We directly compared the model's results to those from children taking part in a number of empirical tasks and showed that, simply by varying the amount of input received by the model, we can successfully simulate a number of effects that we see in the children's performance: individual differences in the rate of vocabulary growth between 19 and 30 months of age; correlations between vocabulary size, familiar word processing speed (as measured by Looking-While-Listening tasks) and phonological working memory capacity (as measured by non-word repetition tasks), and the effect of processing speed and non-word repetition performance on subsequent vocabulary growth. The results suggest that exposure to linguistic input, read through a fixed capacity processing constraint, leads learners to store phonological knowledge in chunks of information of varying size, and it is this stored knowledge that, at least partially, determines familiar word processing speed, non-word repetition performance, and vocabulary growth.

Keywords: inter-individual variation, vocabulary acquisition, CLASSIC model, speed of processing, non-word repetition, working memory capacity

https://doi.org/10.1075/tilar.35.04row
Available under the CC BY-NC-ND 4.0 license.© 2026 John Benjamins Publishing Company

1. Introduction

Explaining why there are individual differences in language acquisition has always been a core goal of usage-based approaches. Usage-based theories start from the premise that acquisition is determined by use, and, in particular, by the form and function of the language that children experience in their everyday lives (Behrens, 2009; cf. Slobin, 2026 this volume). The process of acquiring a language involves extracting regularities from the input via semantic-distributional learning and using these to construct abstract linguistic representations. Learning relies primarily, if not solely, on general cognitive learning mechanisms that perform functions such as chunking (Gobet, Lane, Corker, Cheng, Jones, Oliver, & Pine, 2001; McCauley & Christiansen, 2011), analogy (Behrens, 2017; Gentner, 2010) and schematisation (Behrens, 2021).

On this account, as long as there is variation in the learning environment across children, individual differences are an inevitable consequence of the acquisition process (Kidd, Donnelly, & Christiansen, 2018). For example, Figure 1 shows the differences in the size of the production vocabulary of nearly 5000 American English children, recorded using Communicative Development Inventories (CDIs; parent report checklists in which parents tick off the words their children know and produce). Differences in vocabulary size across children emerge early and grow over time, with some 30-month-olds producing all the words on the 600+ word checklist, but others producing fewer than 50 words. Such differences have been reported in every language studied thus far (Frank, Braginsky, Yurovsky, & Marchman, 2021), and persist throughout childhood, with implications for well-being and educational achievement in later life (Law, Rush, Schoon, & Parsons, 2009). Thus, establishing the causes of individual differences addresses not only an important scientific problem but also a key societal one.

The goal of the present paper was to test a usage-based computational model of inter-individual variation in vocabulary learning, which simulates variation as a product of interactions between linguistic input quantity and the child's current knowledge state. In the remainder of this introduction, we summarise key literature, focussing on the three factors that we simulate in the model, before describing the model's architecture.

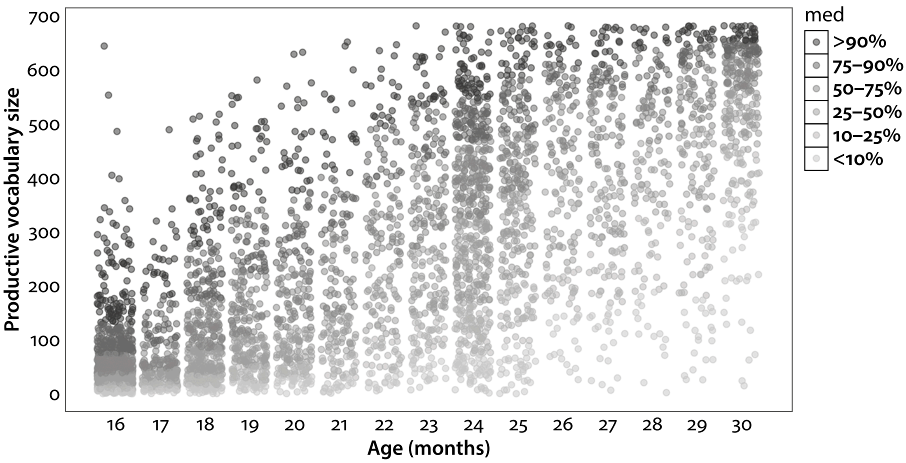


Figure 1. US English-speaking children's productive vocabulary size between 16 and 30 months, as measured by the MacArthur-Bates Communicative Development Inventories (CDI), based on Wordbank data (http://wordbank.stanford.edu/)
Note. Each datapoint represents one child.

1.1 Why do children differ in the speed with which they learn words?

Usage-based approaches have huge potential to explain individual differences, given their fundamental assumption that children construct linguistic representations out of the language that they hear (Behrens, 2009; cf. Slobin, 2026 this volume). From a usage-based approach, differences in the quantity and quality of the input are predicted to play a key role in explaining individual differences, a prediction that is robustly supported by the literature to date (cf. Ellis, 2026 this volume). For example, in a meta-analysis of 52 studies, Anderson, Graham, Prime, Jenkins, and Madigan (2021) reported moderate/large pooled effects sizes for both input quality (defined as vocabulary diversity and syntactic complexity; $r = .33$) and quantity (no. words/tokens/utterances; $r = .20$). Similarly, a recent study of 1001 2- to 48-month-olds across six continents found that input quantity (no. adult words produced in the child's environment) had an equivalent effect size on acquisition (child vocalisation count) as the presence/absence of a clinical disorder (Bergelson, Soderstrom, Schwarz, Rowland, Ramírez-Esparza, Hamrick, Marklund et al., 2023).

However, children are not simple input-output machines; instead they actively construct their linguistic system from the input that they receive (cf. Slobin, 2026 this volume). This means that how, and how quickly, they learn depends on how efficiently they extract and process incoming information (cf. Behrens, 2021). To date, two information processing constraints have been extensively studied: phonological working memory size and familiar word processing speed. Phonological working memory size refers to the amount of phonological information that children can hold in working memory, often measured by how accurately they can repeat non-words of increasing length (non-word repetition (NWR) tasks). It is a processing bottleneck that determines how much material can be processed before the memory trace fades and, thus, constrains the creation of stable phonological representations (and hence words) in long-term memory. Evidence for the relation between phonological working memory size and vocabulary acquisition comes from many studies showing that performance on NWR tasks robustly correlates with vocabulary size both concurrently and longitudinally, across multiple labs, in multiple languages, and in children with and without language disorders (for a review, see Graf Estes, Evans & Else-Quest, 2007; cf. Skoruppa, 2026 this volume).

Processing speed, in the current context, refers to the speed with which children can process a familiar word, usually measured in a Looking-While-Listening (LWL) task and defined as the speed (latency/reaction time) to move the eyes away from a distractor towards a target image when that target is named. Processing speed also acts as a bottleneck on acquisition; faster processing enables faster/more encoding of novel material while it is active during the processing window, leading to better learning (see e.g., Marchman, Ashland, Loi, Munevar, Shannon, Fernald, & Feldman, 2023). Familiar word processing speed in young children is robustly associated with vocabulary size concurrently and longitudinally in studies across multiple labs and languages, and in children with and without developmental delay (Fernald & Marchman, 2012; Hurtado, Marchman, & Fernald, 2007; Marchman, Loi, Adams, Ashland, Fernald, & Feldman, 2018; Peter, Durrant, Jessop, Bidgood, Pine, & Rowland, 2019; Ståhlberg-Forsén, Latva, Leppänen, Lehtonen, & Stolt, 2022).

In sum, extensive work implicates input quantity/quality, working memory capacity, and processing speed as highly inter-related causal factors in the speed of new word learning. However, one important piece of the usage-based puzzle is missing from this literature; the role of the child's own developing knowledge-base. Just as adults process their input using both information from prior experience (top-down information) and information from the incoming signal (bottom-up information; see Hasson, Chen, & Honey, 2015), so too do children (see e.g., Stärk, Kidd, & Frost, 2022). Crucially, in children, the top-down information is constantly changing as a result of new learning, such that each new skill that children acquire will influence how they subsequently perceive, process and, crucially, learn from, their environment. For example, once children learn to interpret eye gaze as a cue to communicative intent, they use this to narrow down a word's possible meaning (Çetinçelik, Rowland, & Snijders, 2020), once they have learned that objects that share a shape tend to share a label, they can apply this 'shape bias' to new object word learning (Perry & Samuelson, 2011), and once they have built some basic semantic/syntactic categories, they can use these to interpret novel verbs in sentences (Gertner, Fisher, & Eisengart, 2006). In other words, as the child's knowledge changes with development, so too does the way in which they process, and learn from, their input (Masten & Cicchetti, 2010).

The prior knowledge that children bring to the learning task also affects information processing efficiency. Children with larger stores of linguistic information in long-term memory are at an advantage, since they can use this stored knowledge to more quickly and efficiently process incoming information. For example, the ability to call on more, and larger, sub-lexical representations from long-term memory to represent the phonological form of non-words facilitates performance on NWR tasks (see, e.g., Jones, Gobet, & Pine, 2007; Szewczyk, Marecka, Chiat, & Wodniecka, 2018). Similarly, vocabulary size affects familiar word processing, such that adults with larger vocabularies show more accurate and faster processing of familiar words (Mainz, Shao, Brysbaert, & Meyer, 2017) and bilingual children's processing speed in one language correlates more strongly with their language experience in that language than their experience in the other (Marchman, Fernald, & Hurtado, 2010). The explanation here is that those with larger vocabularies have not only been exposed to more words, but have also been exposed to each word more frequently (Kuperman & Van Dyke, 2013), which leads to more robust and distinct lexical representations that can be accessed faster during online language processing tasks.

Thus, we know that there are at least three interlocking, inter-dependent factors that affect the speed of vocabulary acquisition: the input that children hear, the constraints imposed by the cognitive mechanisms that children use to process and learn from their input, and the knowledge-base they bring to the language-acquisition task. Our challenge is to come up with an account of how these three interlocking factors lead to individual differences that explains precisely how this developmental process works. To create a usage-based account of inter-individual variation in the speed of vocabulary acquisition, we need a model (a) with a domain-general cognitive learning mechanism, (b) with plausible domain-general constraints on information-processing and learning, and (c) that can be used to simulate the process through which children build vocabulary knowledge out of realistic linguistic input. Preferably this model will also (d) yield output that can be directly compared with that of real children.

In the present paper, we present one such attempt, using a computational model called CLASSIC (*Chunking Lexical and Sub-lexical Sequences in Children*; Jones et al., 2007; Jones, Gobet, Freudenthal, Watson, & Pine, 2014; Jones, 2016). We used this model to determine whether we can simulate inter-individual variation in vocabulary growth as a product of interactions between linguistic input quantity and the child's current knowledge state, using a simple associative learning mechanism with a fixed capacity limited processing window. We compared the output of this model with that of 2-year-old children learning English and assessed its ability to explain the robust relationship between vocabulary size, familiar word processing speed, and performance on NWR tests that we see in the developmental literature.

1.2 Introducing CLASSIC

CLASSIC is a derivative of the EPAM/CHREST architecture (Gobet et al., 2001) and uses an associative chunking mechanism to process incoming utterances and expand the current knowledge base. Chunking is a process where sequences of information (in this case, phonemes) are compressed and recoded into a single perceptual unit (Cowan, 2010; Miller, 1956). Combining separate units of information into chunks enables humans to extract, and remember, longer sequences of information despite limited processing capacities. Chunking as a mechanism was originally proposed by Miller (1956) to explain memory performance, but has since been successfully applied to a number of phenomena including chess experts' memory for chess positions on a board (Gobet & Simon, 1998), why children with developmental language disorder make optional infinitive errors (Freudenthal, Gobet, & Pine, 2024), and how children learn to segment words out of a continuous speech stream (Perruchet, 2019). Chunking is a key mechanism of human learning and processing, explaining both why experience leads to more efficient processing of a stimulus within the confines of a limited processing window (Cowan, 2022), and why experience influences acquisition (Jones & Rowland, 2017; McCauley & Christiansen, 2011).

Figure 2 provides a schematic representation of how CLASSIC processes and learns from input. CLASSIC implements a limited processing window that restricts processing to 4.5 chunks per utterance on average, based both on work suggesting people can process only 4 or fewer chunks at any one time (Gobet & Clarkson, 2004) and work suggesting that this limit might be slightly larger (Miller, 1956). Within this window, CLASSIC processes as much of an incoming utterance as it can, coding it into as few chunks as possible using whatever chunks it has available in long-term memory (Processing layer, Figure 2), and learns new chunks by grouping adjacent chunks in the processing layer to form a new chunk (Learning layer, Figure 2). Learning is the formation of new chunks that pair adjacent chunks. Crucially these new (longer) chunks immediately become available to the processing layer, to be used in parsing the next incoming utterance. For example, if the words *where's*, *the* and *dog* exist as chunks in long term memory, the phrase *where's the dog* can be encoded as three chunks even though it is nine phonemes long.

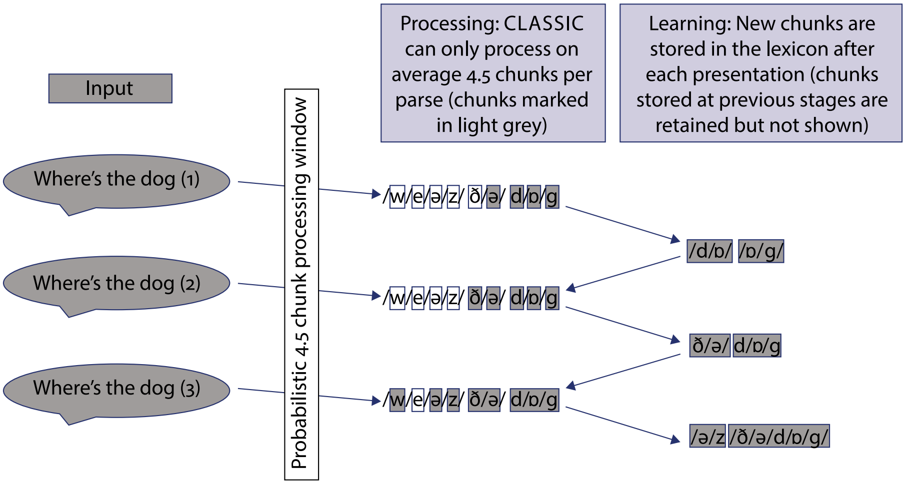


Figure 2. Learning in CLASSIC after three presentations of "Where's the dog"
Note. At the processing stage, grey squares indicate the chunks that CLASSIC has accessed in its processing window (on average 4.5 chunks per parse) which probabilistically favour the ends of utterances. At the learning stage, light grey squares represent chunks learned from the input after each presentation. CLASSIC starts with only knowledge of phonemes. 1st presentation: Processing: CLASSIC is limited to processing 4 chunks. Since it starts out with only knowledge of phonemes, it parses the input as 9 one-phoneme chunks, only 4 of which are accessed for learning ([ə],[d],[ɒ] and [g]). Learning: From the 4 one-phoneme chunks it has accessed, it creates (learns) two new, bigger chunks by combining adjacent accessed chunks and storing them ([d/ɒ/] and [ɒ/g/]; note that CLASSIC does not chunk phonemes across different words unless the words themselves are full chunks. 2nd presentation: Processing: CLASSIC is limited to processing 4 chunks. It has already chunked [d/ɒ/], which enables it to process this two phoneme sequence as one chunk. This means that it can process more of the utterance than it did at the 1st presentation (5 phonemes as 4 chunks: [ð], [ə], [d/ɒ/], and [/g/]). Learning: From the 4 chunks it has accessed, it again combines adjacent chunks, resulting in two new, bigger, chunks: [ð/ə/] and [d/ɒ/g/]. Note that CLASSIC has now learned two complete words, each represented as 1 chunk: the ([ð/ə/]) and dog ([d/ɒ/ g/]). 3rd presentation: Processing: CLASSIC is limited to processing 5 chunks: the newly learned words [ð/ə/] and [d/ɒ/g/] together with 3 more chunks [w], [ə] and [z]. Learning: Adjacent chunks are then chunked again: [ə/z], [ð/ə/ d/ɒ/g/]. Note that because [ð/ə/] and [d/ɒ/g/] are whole words (the, dog) CLASSIC chunks them into a phrase. This image is licensed under CC BY 4.0 https://creativecommons.org/licenses/by /4.0/deed.en

There are two consequences of this cycle. First, as CLASSIC receives more input, it stores more, and longer, chunks of information in long-term memory. Thus, experience with the language leads to CLASSIC learning a large number of sub-lexical, word-level (i.e. lexical), and phrase-level chunks. Second, since these stored chunks can be used to more quickly and efficiently process incoming information, more knowledge means that more input can be processed in the model's limited capacity processing window. The more language CLASSIC is exposed to, the more language it knows (in terms of chunks of sub-lexical and lexical knowledge), the better it performs.

CLASSIC is a simple model, without semantic or syntactic knowledge, which learns the phonological form of words by binding adjacent sequences in the input and learning these sequences as larger chunks. Despite its simplicity, it is surprisingly successful at simulating vocabulary acquisition. For example, it has been used to simulate the effect of input quantity and quality on vocabulary acquisition (Jones & Rowland, 2017) and of individual and developmental differences in performance in a number of NWR tasks (see e.g., Jones, 2016), demonstrating that models that receive more input have larger stores of sub-lexical and lexical chunks, which they can use to both repeat longer non-words in non-word repetition tasks and build vocabulary more quickly.

CLASSIC has not yet been used to simulate the relationship between vocabulary size and familiar word processing speed, but it can be extended to do so. Jones (2012) showed that CLASSIC could simulate developmental increases in processing speed in children simply as a consequence of the number of chunks known by the model. Since the time needed to encode a chunk is roughly the same no matter the length of the chunk (Simon, 1974; Zhang & Simon, 1985), words that can be represented as fewer chunks will be processed faster. By extension, CLASSIC should be able to simulate the relationship between processing speed and vocabulary size simply because models with more word-level chunks in their long-term memory are likely to be faster in the Looking-While-Listening task. For example, a child/model who can represent the word *dog* as one chunk will be faster to respond to the instruction *look at the dog* than a child/model who needs three phoneme-level chunks to represent the same word.

Defining the relationship between vocabulary knowledge, working memory capacity and processing speed in terms of chunked knowledge has two additional consequences. First, it means that processing speed and working memory (or processing) capacity should be intrinsically related concepts; a child with more and bigger chunks in their long-term memory will both process familiar words faster in speed of processing tasks and be able to retain, and reproduce, more non-words during NWR tasks (Case, Kurland, & Goldberg, 1982). Thus, we should expect a correlation between performance on NWR tasks and familiar word processing speed in both CLASSIC and children. Second, since input needs to be processed before it can be learned, both processing capacity and speed should affect subsequent vocabulary growth (i.e., determine how much information from an incoming utterance can be learned). In other words, performance on NWR and processing speed tasks should predict subsequent vocabulary growth in both CLASSIC models and children. At this point it is worth noting that processing capacity is fixed in CLASSIC, but that more information can be processed as learning progresses. Speed of processing is also determined by what knowledge CLASSIC has at its disposal. In essence, CLASSIC suggests that perceived differences in processing capacity and speed of processing arise from an increasing knowledge-base.

1.3 The present study

In the present study, we determined whether CLASSIC can simulate inter-individual variation in vocabulary growth between 19 and 30 months of age, and its relation with performance on NWR and processing speed tasks. The advantage of implementing the usage-based theory in a computational model is that modelling requires us to be precise in our descriptions of learning mechanisms; without such implementation, it is never entirely clear if such mechanisms are sufficient or necessary to account for the data. A more specific advantage of CLASSIC is that it learns from real language, which means that we could directly compare the model's output with data from a large cohort of children who took part in speed of processing (at 19 months) and NWR tasks (at 25 months) and whose vocabulary size was measured at multiple time-points between 19 and 30 months of age.[1]

We compared the performance of the children and models to determine whether, by varying the amount of input received by the model, we could successfully simulate the following: (1) individual differences in the rate of vocabulary growth between 19 and 30 months of age, correlations (2) between vocabulary size and a NWR task at 25 months and (3) between vocabulary size and performance on a processing speed task at 19 months, and (4) the effect of processing speed and NWR performance on vocabulary growth over the next 6–12 months. We also (5) tested for a correlation between LWL and NWR performance in both children and models – a prediction derived from our hypothesis that both processing speed and working memory capacity are at least partly determined by the number and size of chunks of information in long-term memory. In the interests of

1. A point to note, however, is that CLASSIC only simulates the acquisition of the phonological form of words, not the link between form and meaning (see discussion section for more on this).

full disclosure, note that some analyses have also been published elsewhere (Peter et al., 2019; Rowland, Bidgood, Jones, Jessop, Stinson, Pine, Durrant et al., 2024).

2. Method

2.1 Participants

This study is part of a longitudinal cohort study that tracked the language development of children between six months and 4;6 years of age (the Language 0–5 Project, see Rowland, Durrant, Peter, Bidgood, Pine, & Jago, 2018). Ninety-five monolingual British English-speaking families were recruited, one child was excluded because of a persistent ear infection, and four families did not continue after the initial visit. By the end of data collection (at age 4;6), a further 13 had dropped out.

LWL tasks to assess familiar word processing speed were administered at the 19-, 25- and 31-month datapoints. Because there was no relationship between processing speed and vocabulary at the 25- and 31-month datapoints, we do not include those results here (see Peter et al., 2019, for full results). At the 19-month datapoint, 80 children (42 females) were tested, 76 of whom provided usable data (had more than 25% good trials; mean age 19;3 months, range = 19;1–20;1). NWR tasks to assess phonological working memory capacity were administered at the 25- and 31-month age-points. Because the results from the 25- and 31-month datapoints yielded very similar patterns, we focus on the 25-month datapoint only (see Rowland et al., 2025, for full results). At 25 months, 75 children took part in the NWR task, 67 of whom (35 females) provided usable data (i.e., at least one valid response; mean age 25.86, range = 25.17–26.73). At the time of testing, all of the children were typically-developing, all were monolingual English learners, and none had any reported problems with vision or hearing.

2.2 Stimuli and procedure

Vocabulary size. Vocabulary size was assessed at 3-month intervals from 9 to 36 months of age using parent report checklists (Communicative Development Inventories) created for British English. For the present study, we used data from the Lincoln CDI (Meints, Fletcher, & Just, 2017), adapted from the MacArthur-Bates CDI Words & Sentences (Fenson, Marchman, Thal, Dale, Reznick, & Bates, 2007), which was administered at the 19-, 21-, 24-, 25-, 27-, and 30-month datapoints. This is a parent report checklist that contains, among other sections, a vocabulary scale of the most common vocabulary items in UK children's vocabulary between 18 and 30 months of age (total possible score = 689).

Non-word repetition (NWR) task. Full details can be found in Rowland et al. (2024). A new NWR task for 2-year-olds was created, comprising 36 non-words divided into two lists of 18 non-words that were assigned to the 25-month and 31-month age-point respectively. Each list contained six one-syllable, six two-syllable and six three-syllable non-words, half of which were wordlike and half not wordlike. The non-words were placed in semi-random order within the lists (1-syllable – 2-syllable – 3-syllable).

The task was embedded in a Fuzzy Felt game in which a puppet would say some "funny-sounding words" and the child's task was to repeat them. Each time the child attempted to copy a non-word, they received a Fuzzy Felt sticker to place on a board, to build up a full picture by the end of the session. The session began with three real word practice items (*cow*, *button*, *elephant*) followed by the eighteen non-words. All non-words were produced live by the experimenter using a natural prosodic pattern (maximum two times), always in a carrier phrase (e.g., *Can you say …?*). If the child failed to respond the second time, the experimenter encouraged the child once more but without repeating the word (e.g., *Can you say it? What did he say?*). If there was still no response, they moved onto the next item.

Looking-While-Listening (LWL) task. Full details can be found in Peter et al. (2019). The task replicated as closely as possible the task used by the original authors (e.g., Fernald, Perfors, & Marchman, 2006). All sentences were pre-recorded by the same female native English speaker and normalised for pitch and volume. Each sentence included a two-part question and the target word (e.g., *Where's the baby? Can you find it?*), with varied questions to keep the child engaged. Sentences were presented in two blocks (A and B) so that target words in block A were distractor items in block B. All children were presented with both blocks, but block order was counterbalanced across children so that half were presented with block A first and the other half with block B first. There were 64 test trials across two blocks: eight different nouns (e.g., *baby*) each presented eight times as a target noun. Each block also included three filler sentences (e.g., *Well done! Do you like the pictures?*). The visual stimuli were pictures of objects that matched the chosen target words shared by the original authors.

Children's eye movements were coded online using an eye-tracker (EyeLink 1000 Plus; SR Research: Ottawa, Ontario, Canada). The child sat in a car seat, high chair or on their caregiver's lap in front of a 17-inch LCD monitor mounted on a hydraulic arm. The experimenter first played a short cartoon to familiarise the child with the set-up and to orient the child's eyes to the screen, then calibrated using a 5-point callibration sequence in which the child saw a looming high-contrast circular shape accompanied by a twinkly sound. The experiment began with the presentation of a central smiley face on the screen, that acted as a gaze-contingent attention getter; the next trial began after the child had fixated on the face for 8ms. For every test trial, the child first saw two images – the target and the distractor – presented simultaneously on the left and right side of the screen. These images were presented in 2000ms of silence before the onset of the speech. Both images remained on the screen for the entirety of the trial (7000ms).

2.3 CLASSIC training

Full details are in Rowland et al. (2025). For this study, CLASSIC was re-built using Python 3.10. As input, fully intelligible caregiver utterances from English UK corpora from the CHILDES database (MacWhinney, 2000) were extracted for ages 1;10 to 4;10 using the package childesr (Braginsky, Yurovsky, Marchman, & Frank, 2019). This resulted in 712,441 utterances directed to 2–3-year-olds, 294,670 utterances directed to 3–4-year-olds and 67,130 utterances directed to 4–5-year-olds. Each word in the transcripts was converted to its constituent phonemes using an English lexicon (https://github.com/cmusphinx/cmudict) with phonemic transcripts for unknown words added manually for all words occurring with a frequency of 100 or more across all transcripts. This enabled us to retain 96% of the original orthographic utterances in their phonemic form. The boundaries between words were also retained.

The number of utterances presented to CLASSIC was manipulated to capture individual differences in the amount of input that children hear. Data from transcripts were pooled, and novel input samples were generated by randomly selecting utterances from the pooled data. Temporal characteristics of the utterances were retained by creating 10 sample bins from which we randomly selected utterances, each reflecting the time-point at which the utterance occurred (e.g., bin 1 = utterances to 2–3-year-olds etc.).

We created 80 unique input quantity levels, ranging from small (1500 utterances) to large (120,000), with each level representing an increase in increments of 1500 utterances. Since learning in the model is dependent on a probabilistic processing constraint (i.e., while on average, 4.5 chunks will be accessed for every utterance, this will vary across utterances), the results were slightly different each time the models were run. Thus, five separate samples were generated for each quantity level and were individually presented to the model. In total, the results of the present work were based on 400 (5 x 80 input levels) models.

2.4 Comparing children and models

We estimated the stage in the learning cycle that the models could be considered to be equivalent to a 19- and 25-month-old child in terms of CDI vocabulary size using the median number of words known by English children according to the CDI data on Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2016). To match to the 19-month age-point, we extracted from Wordbank the median number of words that were produced by 19-month-old English-learning children (median = 104). We then identified the point in the learning cycle at which the median model also 'knew' approximately 104 words on the CDI, which was after 5% of input had been seen (see below for how we calculated how many words the models knew). We used this point in the learning cycle (5%) as the 19-month age-point for all models. To match to the 25- and 30-month age-points, we extracted from Wordbank the median number of words that were produced by 25- and 30-month-old English-learning children (373 and 558 words respectively) and identified the point in the learning cycle at which the median model also knew approximately this number of words on the CDI (13% of the input seen for 25-month, and 48% for 30-month datapoint).

2.5 Coding

Vocabulary size. For the children, expressive vocabulary CDI scores were calculated at the 19-, 21-, 24-, 25-, 27- and 30-month datapoints according to the instructions in the CDI manual. For the models, CDI scores were calculated at regular intervals during training by identifying how many words from the Lincoln CDI were represented as a single chunk in the model's long-term memory (i.e., had been learned as one whole entity; this included frozen phrases such as *thank you*). Where the CDI offered alternatives (e.g., *telly*, *TV* or *television*), knowledge of any of the items as a single chunk was taken as knowledge of the vocabulary item.

Non-word repetition (NWR) task. The children's repetitions of the non-words were phonemically transcribed in ELAN and exported into a csv file for coding. The following were excluded: null responses, unclear responses (e.g., mumbled responses), responses to targets in which the experimenter's voice was obscured (e.g., by background noise), responses in which the experimenter made a pronunciation error or exceeded the number of presentations allowed, items repeated spontaneously by the child (e.g., not immediately preceded by the target), and items for which the parent prompted the child. If a child repeated a non-word, we scored their first scorable attempt. There were, on average, 16.43 valid responses per child out of 18 possible responses (range = 8–18, SD = 2.56).

Non-words were coded as correct or incorrect (at least one error) but articulation errors that are common in the speech of 2-year-olds were allowed (see Rowland et al., 2025, for more details). Inter-rater agreement, calculated from the data from eight participants (items = 276) by two coders trained in phonemic transcription, was 93.8% (Cohen's Kappa = 0.86). Each participant's total NWR score was calculated as a proportion of the total number of valid responses to yield a total proportional score out of 1.

The models were given the same non-words as the children at the 25-month datapoint (i.e., after 13% of training). No learning took place during the test battery. Each non-word was presented 100 times to gain a more reliable estimate given the probabilistic processing constraint. The outcome variable for each item was the proportion of presentations in which all the chunks needed to process the non-word fit within the 4.5-chunk processing constraint (i.e., the proportion of times (out of 100) that the non-word could be considered to have been 'repeated' correctly), which was then averaged across the five models at each input level. The models' total NWR score was calculated by averaging across the scores for each non-word to yield a total proportional score out of 1.

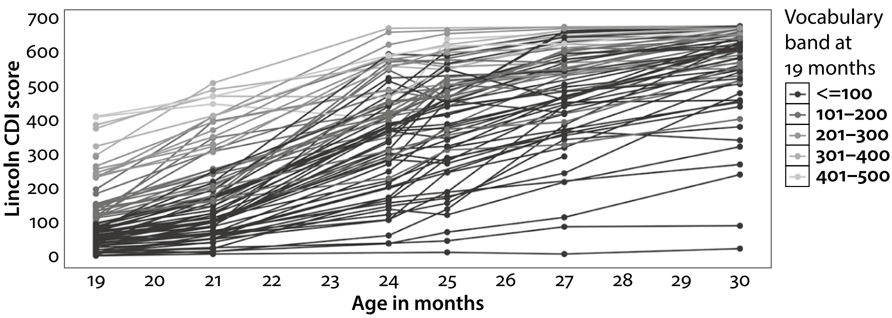
Looking while listening (LWL) task. The eye-tracker recorded the children's eye movements, with interest areas set to the dimensions of the target and distractor images. Speed of processing was defined as the time that it took for the child to initiate a shift from the distractor image to the target image (reaction time (RT)), and included any shifts that occurred in the time-window from 300 to 1800 ms after target word onset. Trials were excluded if there was audible disruption (e.g., crying, talking) at target word onset, the target word was not understood by the child, as reported by the caregiver, or the child did not fixate towards both the distractor and target during the pre-speech period. In addition, RT was only calculated on trials in which the child was looking at the distractor at the onset of the target word (distractor-initial trials), since speed of processing is defined as the initiation of a shift *from* the distractor *to* the target image.

The models were given the same LWL items as the children at the 19-month datapoint (after 5% of training input). No learning took place during the test battery. For each item, processing speed was defined as the number of chunks needed to process the target utterance, on the assumption that words that can be processed in fewer chunks will be processed faster. As with the NWR task, the LWL score was averaged across 100 presentations.

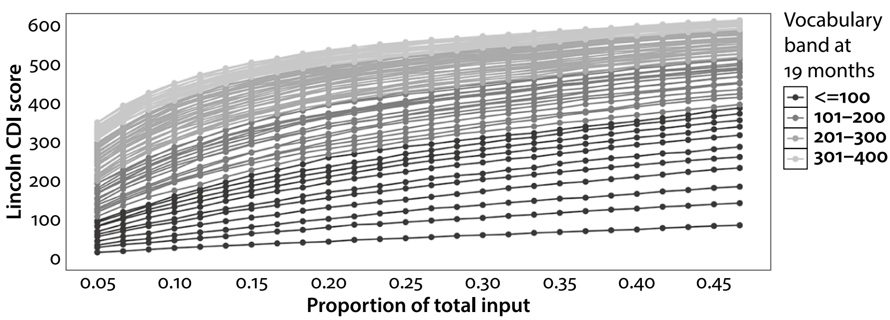
3. Results

3.1 Vocabulary growth

We used R Core Team (2024) for all analyses. Figure 3 shows the growth of CDI vocabulary in children and models. Although the trajectory of vocabulary growth is much noisier in the children, CLASSIC does simulate individual differences in vocabulary growth between 19 and 30 months of age. Note that, in the model, we know that differences in vocabulary size stem from differences in the size of the input sample given to the models. This can be seen in Figure 4, which shows the models' 19-month vocabulary score plotted against total input size. Models that had received more input by the 19-month datapoint had bigger vocabularies.



a. Children



b. Models

Figure 3. Productive vocabulary size between 19 and 30 months for (a) children and (b) models. *Note*. Each line represents one child/model.

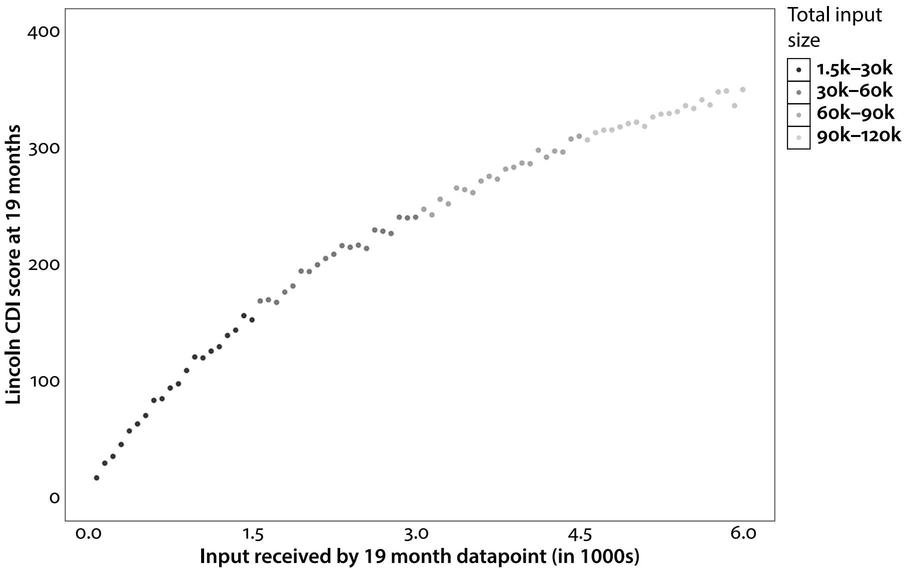


Figure 4. Models: Relationship between input sample size and vocabulary size at 19 months (after 5% of the input)

3.2 NWR tasks and vocabulary size

We fitted Pearson's correlations with bootstrapped confidence intervals to the children and model's performance on the NWR task and vocabulary size at 25 months. As predicted, there were significant, positive correlations for children (Pearson's $r = 0.51$ [0.34, 0.70], $df = 60$, $p < .001$) and models (Pearson's $r = 0.93$ [0.90, 0.96], $df = 78$, $p < .001$; all ps one-tailed since the predictions are directional). Children and models with higher NWR scores had higher CDI vocabulary scores at 25 months of age.

3.3 Processing speed task (LWL) and vocabulary size

We fitted Pearson's correlations with bootstrapped confidence intervals to the models' 19-month LWL and CDI vocabulary scores and compared them to those from the children reported in Peter et al. (2019). As predicted there were significant, negative correlations for children (Pearson's $r = -0.40$, $p < .001$, reported in Peter et al. (2019)) and models (Pearson's r [95% CI] = −.939 [−.959, −.913], $df = 78$, $p < .001$; all ps one-tailed). Children and models with faster processing speeds had higher CDI vocabulary scores at 19 months of age.

3.4 Processing speed (LWL) and NWR tasks

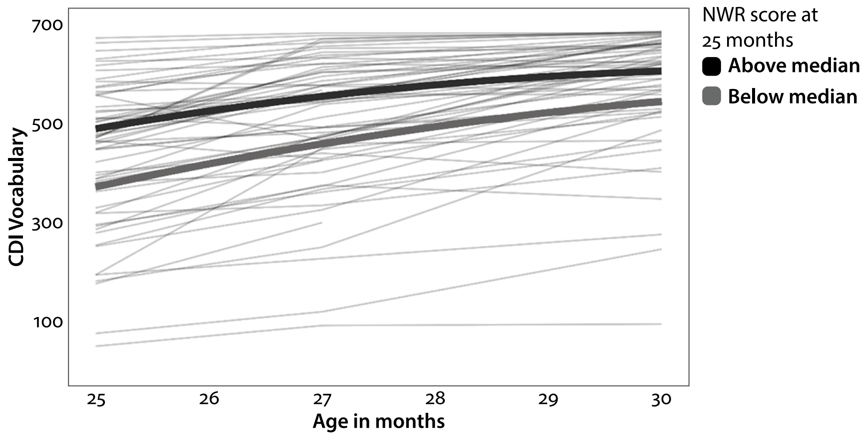
If familiar word processing speed and NWR performance are both at least partially determined by the amount of chunked sub-lexical and lexical material stored in long-term memory, we predict a strong, negative, significant correlation between LWL and NWR scores. We fitted Pearson's correlations with bootstrapped confidence intervals to the children's and models' LWL and NWR scores. As predicted, there were significant, negative correlations for children; Pearson's r [95% CI] = −.349 [−.606, −.085], $df = 62$, $p = .028$, (all ps one tailed) and models; Pearson's r [95% CI] = −.941 [−.983, −.907], $df = 78$, $p < .001$). Children and models with faster processing speeds also had higher NWR scores.

3.5 Predicting vocabulary growth from NWR performance

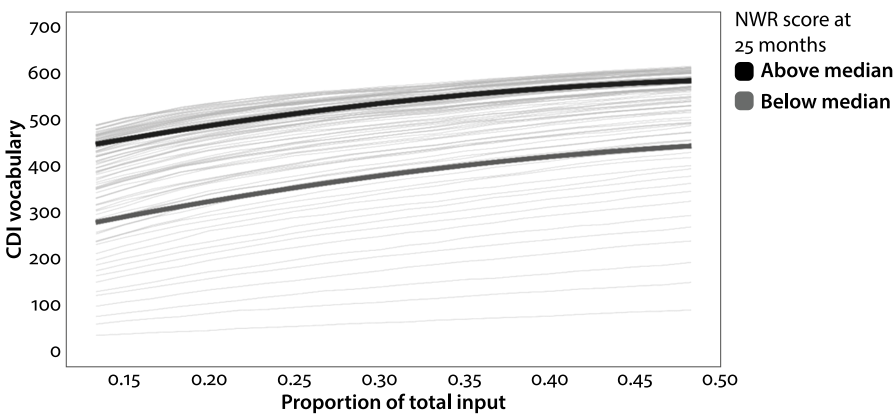
Next we determined whether NWR performance at 25 months of age predicted later vocabulary growth in children and models. We fitted growth curve models where the outcome measure was vocabulary score between 25 and 30 months (Lincoln CDI) and development was modeled using linear (Age1) and quadratic (Age2) effects of age/training (not centred or standardardised).[2] NWR score was entered as a continuous predictor, crossed with both polynomial age predictors. For the models, the maximal random-effects structure supported by the data included subject as a random intercept with a by-subject random slope for Age1 and Age2, their interaction and the random effects correlation parameter. For the children, the maximal random-effects structure supported by the data included subject as a random intercept, with Age1 and Age2 as random slopes, but with no correlation parameter between random effects. The results are illustrated in Figure 5 and Tables 2 (children) and 3 (models).

For both children and models, there were main effects of linear and quadratic age on vocabulary growth, indicating that vocabulary grew with age, and a significant effect of NWR performance, indicating that children/models with higher NWR scores had larger vocabularies over the next six months. There was a significant interaction between NWR performance and linear growth but this was negative, indicating that children/models with better NWR performance showed decelerating, rather than accelerating, vocabulary growth over time. This is almost certainly because the fastest developing children/models reached ceiling on the CDIs before the end of the testing period.

2. The linear term tells us how well vocabulary growth is explained in terms of linear growth (i.e. a consistent rate of growth over time). However, vocabulary growth is rarely linear so we also used a quadratic term, which captures curvilinear growth (e.g. whether the speed of vocabulary growth changes over the period tested)



a) Children



b) Models

Figure 5. Relationship between NWR scores at 25 months and subsequent CDI vocabulary growth in (a) children, (b) models

Note. The thick lines illustrate language growth for children/models with NWR scores below and above the median at the 25 month datapoint. Thin lines show the developmental trajectories of individual children/models. A version of this figure can also be found in Rowland et al. (2025).

Table 2. Children: Result of models estimating effect of NWR performance on subsequent vocabulary growth
Note. Confidence intervals computed using parametric bootstrapping ($r = 1000$). P-values obtained via t-test with Satterthwaite's method.

Term	b	SE	df	t	p	95% CI
(Intercept)	508.63	13.79	64.14	36.89	<.001***	[480.79, 535.53]
Linear age	762.71	57.29	62.51	13.31	<.001***	[651.73, 874.70]
Quadratic age	−95.89	36.13	61.12	−2.65	.010*	[−165.93, −22.94]
NWR	56.62	13.72	64.24	4.13	<.001***	[28.42, 83.73]
NWR x linear age	−187.12	57.76	63.76	−3.24	.002**	[−293.06, −66.71]
NWR x quadratic age	0.74	36.33	62.64	0.02	.984	[−71.39, 71.39]

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Fit statistics

AIC	AICc	BIC	R2_conditional	R2_marginal	ICC	RMSE	Sigma
2,103.50	2,104.75	2,135.82	0.94	0.34	0.91	23.52	35.32

Table 3. Models: Result of models estimating effect of NWR performance on subsequent vocabulary growth
Note. Confidence intervals computed using parametric bootstrapping ($r = 1000$). P-values obtained via t-test with Satterthwaite's method.

Term	b	SE	df	t	p	95% CI
(Intercept)	457.87	3.35	78.00	136.51	<.001***	[450.85, 464.66]
Linear age	315.98	14.04	78.00	22.51	<.001***	[291.16, 342.64]
Quadratic age	−418.66	9.36	78.00	−44.72	<.001***	[−436.59, −399.12]
NWR	99.21	3.35	78.00	29.57	<.001***	[92.92, 105.46]
NWR x linear age	−105.41	14.04	78.00	−7.51	<.001***	[−135.27, −77.23]
NWR x quadratic age	−63.11	9.36	78.00	−6.74	<.001***	[−80.59, −46.81]

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Fit statistics

AIC	AICc	BIC	R2_conditional	R2_marginal	ICC	RMSE	Sigma
10,148.28	10,148.48	10,219.43	1.00	0.92	0.99	3.03	3.22

3.6 Predicting vocabulary growth from speed of processing

The final analyses determined whether LWL performance at 19 months of age predicted later growth in vocabulary in the models and compared it to the results reported for the children in Peter et al. (2019). Because growth was again not linear, we fitted growth curve mixed effects models. The outcome measure was vocabulary scores between 19 and 30 months (Lincoln CDI). Development was modelled using linear (Age1) and quadratic (Age2) effects of age/training (not centred or standardardised). LWL scores were entered as a continuous predictor, crossed with both polynomial age predictors. The maximal random-effects structure supported by the data included subject as a random intercept with a by-subject random slope for Age1 and Age2, their interaction and the random effects correlation parameter. The results are illustrated in Figure 6 for the children (reproduced from Peter et al. 2019) and Figure 7 for the models. Parameter estimates and fit metrics are in Tables 4 (children) and 5 (models).

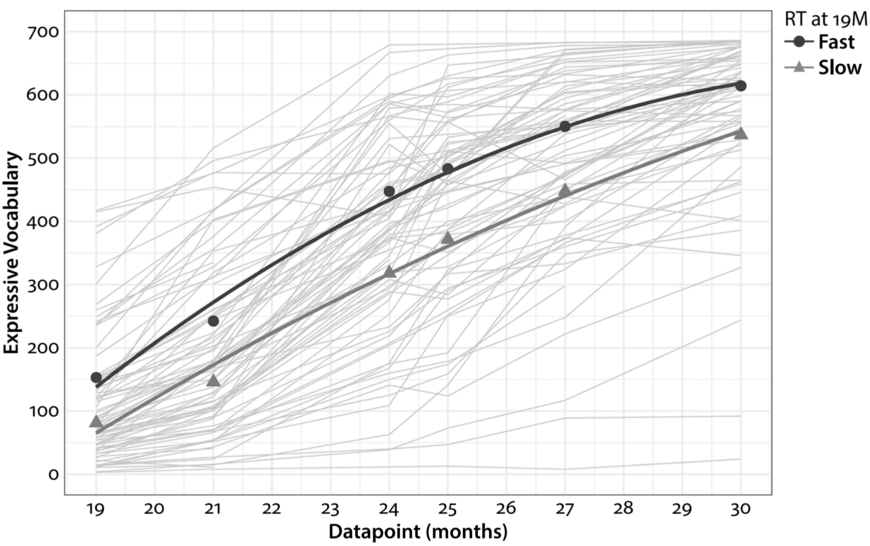


Figure 6. Children: Effect of Looking-While-Listening performance at 19 months on subsequent vocabulary growth
Note. Thick lines: language growth for children with LWL scores below/above the median. Thin lines: developmental trajectories of individual children. RT = reaction time. Figure reproduced from Peter et al. (2019) with permission.

For both children and models, there was a main effect of linear and quadratic age on vocabulary growth, indicating that vocabulary grew with age. Crucially there were significant effects of LWL performance, indicating that children/mod-

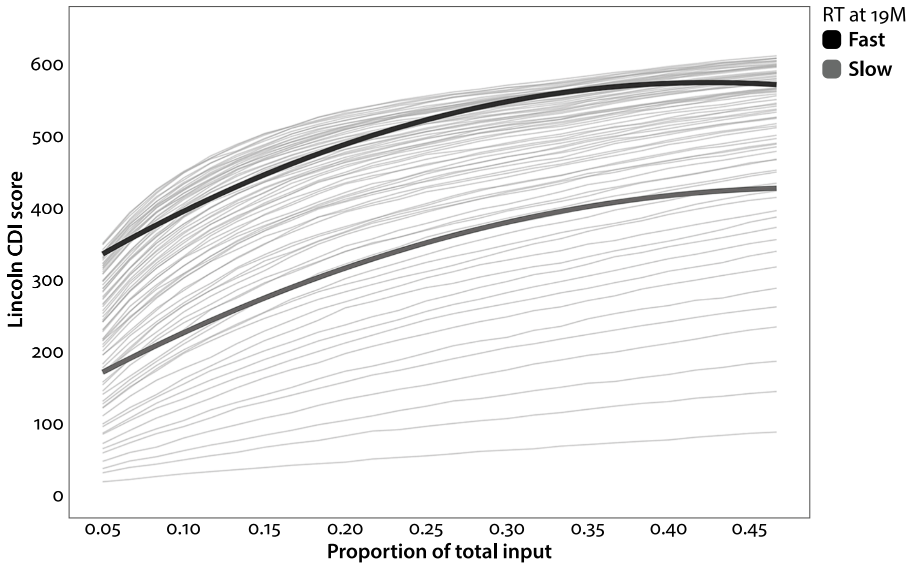


Figure 7. Models: Effect of Looking-While-Listening performance at 19 months on subsequent vocabulary growth
Note. Thick lines: language growth for models with LWL scores below/above the median. Thin lines: developmental trajectories of individual models. RT = reaction time.

Table 4. Children. Result of models estimating effect of LWL performance on subsequent vocabulary growth
Note. Confidence intervals computed using parametric bootstrapping ($r=1000$). P-values obtained via t-test with Satterthwaite's method. Adapted from Peter et al. (2019) with permission.

Term	*b*	*SE*	*t*	*p*	95% CI
(Intercept)	362.68	12.68	28.61	–	[337.74, 387.44]
SoP	−0.39	0.09	−4.22	<.001***	−0.56, −0.21]
Linear age	385.28	13.31	28.94	<.001***	[358.67, 410.86]
Quadratic age	−56.64	9.45	−5.99	<.001***	[−75.28, −38.23]
SoP x linear age	0.02	0.09	0.26	.583	[−0.16, 0.21]
SoP x quadratic age	0.16	0.07	2.26	.286	[0.02, 0.29]
AIC	BIC	logLik	Deviance		
5106.3	5159.4	−2540.1	5080.3		

Table 5. Models: Result of models estimating effect of LWL performance on subsequent vocabulary growth
Note. Confidence intervals computed using parametric bootstrapping ($r = 1000$). P-values obtained via t-test with Satterthwaite's method.

Term	*b*	*SE*	*df*	*t*	*p*	95% CI
(Intercept)	362.24	1.59	78.00	228.03	<.001***	[358.63, 364.81]
SoP	−69.66	1.59	78.00	−43.84	<.001***	[−73.79, −67.13]
Linear age	471.19	13.25	78.00	35.57	<.001***	[459.51, 492.32]
Quadratic age	−462.43	10.13	78.00	−45.65	<.001***	[−471.56, −445.46]
SoP x linear age	53.90	13.25	78.00	4.07	<.001***	[34.73, 87.19]
SoP x quadratic age	67.92	10.13	78.00	6.70	<.001***	[55.29, 93.62]

Note. * $p<.05$, ** $p<.01$, *** $p<.001$

Fit statistics

AIC	AICc	BIC	R2_conditional	R2_marginal	ICC	RMSE	Sigma
12,966.52	12,966.81	13,062.40	1.00	0.24	1.00	3.70	3.93

els with faster processing speeds had larger vocabularies, and that this difference was sustained over the next 12 months. For the models, there was a significant interaction between NWR performance and linear and quadratic growth, but this was because models with better LWL performance showed decelerating, rather than accelerating, vocabulary growth over time. From the graph, we can see that the children show a similar, but less pronounced, effect.

4. Discussion

Our goal was to determine whether we could simulate inter-individual variation in vocabulary growth as a product of interactions between linguistic input quantity and the child's current knowledge state. We also tested if we could simulate the relationship between vocabulary size and working memory capacity (measured by the NWR task) and familiar word processing speed (measured by the LWL task).

We used the computational model CLASSIC that implements three aspects of the usage-based thesis: (1) it learns from real input to children (text-based, phonetically coded British English utterances spoken by caregivers to children during play); (2) it learns using an associative, chunk-based learning mechanism with a limited, fixed capacity processing window, and (3) new learning is influenced not just by input but by the amount and type of knowledge stored in the lexicon (as lexical and sub-lexical chunks of linguistic information).

We successfully simulated individual differences in growth of vocabulary between 19 and 30 months of age, differences that, in the models, resulted solely from differences in the amount of input that the models received. In both children and models, there were already substantial differences in vocabulary size at 19 months of age (after 5% of the input to the model), and these differences were maintained up until 30 months of age (after 48% of the input). There were also significant positive correlations between NWR performance and vocabulary; children who could accurately repeat longer non-words, and models who could represent non-words in fewer chunks, had bigger vocabularies. Similarly, there were significant negative correlations between LWL performance and vocabulary size and also between LWL and NWR performance. In the models, this was to be expected, because familiar word processing speed and NWR performance are both determined by the amount of chunked sub-lexical and lexical material stored in long-term memory. In the children, this finding is more substantial, as it indicates that processing speed and processing/working memory capacity are linked, both influenced by the amount and type of knowledge stored in the lexicon. However, in all cases, the correlations were stronger in the models than in the children; a point that we return to below.

Finally, we demonstrated that the relationship between vocabulary size and LWL and NWR performance persisted over time. In both children and models, both LWL and NWR performance not only correlated with concurrent vocabulary but also predicted later vocabulary scores. Thus, there were long-term relationships between NWR performance, LWL performance, and vocabulary in 2-year-olds, relationships which we successfully simulated in the CLASSIC model.

Importantly, we did not build into CLASSIC any intrinsic differences in either processing speed or working memory capacity. Thus, the results suggest that, in line with usage-based approaches, it is possible to capture a substantial proportion of inter-individual variation in vocabulary acquisition as a product of interactions between linguistic input quantity and the child's current knowledge state, using a simple associative learning mechanism. On this view, children exposed to more language input learn more linguistic information – both more words, and more sub-lexical representations – and this stored knowledge speeds up the recognition of familiar words, improves performance the NWR task, and facilitates learning new words.

That said, there were key differences in how the models and children performed. The trajectory of vocabulary growth in the children was much noisier than that of the models, and some children even had lower scores at later datapoints than at earlier ones. This is partly because the data comes from parent report instruments, which are reliant on parents accurately recalling the words their children know, and partly because children can forget previously learned words, whereas the models cannot. In addition, although vocabulary growth in the faster children and in the faster models started to asymptote before the end of the learning period, for the children, this was because they had learned all 689 words on the checklist, but for the models, this occurred before they had reached the 600 word mark. This difference is almost certainly due to the restricted nature of the models' input compared to 2-year-old children who will have heard a much wider range of word types over their 2-year lifespan. Some of the words on the CDI simply did not occur in the input sample given to the models. This can be solved by increasing the diversity of the input given to models.

Another notable difference was that the correlations between vocabulary size, NWR and LWL performance were consistently smaller in the children than in the models. This is because models simulate a simplified learning process. For example, the model begins with perfect phonological representations of English phonemes, such that knowledge of a chunk equates to a perfect reproduction of the chunk contents, whereas 2-year-olds are still constructing their phonological representations (Snowling, Chiat, & Hulme, 1991) and developing proficiency in articulation (Roy & Chiat, 2004). Similarly, children's performance will be affected by factors not captured in the model (time it takes to plan a saccade, boredom, fatigue etc.). Another difference is in how we operationalised processing speed in the model. Our definition (number of chunks that the model needed to code the word on the basis that words that can be represented in fewer chunks will be processed faster) is a simplification of the relationship between chunk size and processing speed. For example, Zhang and Simon (1985) estimated that it takes approximately 400ms to process and articulate one chunk, but also found that an extra 90ms processing time per syllable was needed after the first syllable, and a previous version of CLASSIC (EPAM-VOC) implemented this (Jones 2011). A similar model, CIPAL (Jessop, Pine, & Gobet, 2025) set timing parameters to determine how long a processing operation took, controlling speech rate (160ms), phonological decay rate (800ms), and the initial retrieval time for new phonemic chunks when they were added to long-term memory (1200ms). The chunks in long-term memory also had individual retrieval times that got faster each time they were retrieved to code the input. Both these models can simulate individual and developmental differences in processing speed, although neither directly compared the models' output directly with those of children on the same task. Future work should explore the implications of different methods of operationalising processing speeds for chunks of varying sizes and familiarity.

Another reason for differences in the output of models and children is the fact that CLASSIC is a very simple model that learns the phonological form of words only, without semantic or syntactic knowledge. Learning is affected by the frequency with which both the word itself, and the phonological material contained in the word, occur in the input, both of which also have an effect on children's word learning (Braginsky et al., 2019). However, in children, word learning is also affected by a number of other factors such as the word's semantic density (the number of words the child knows that have a similar meaning; Borovsky, 2022), concreteness (the extent to which the word denotes a concept that can be experienced by the senses), valence (emotional affect), and even babiness (how much the word is associated with infancy; see for example Braginsky et al., 2019, for evidence pertaining to concreteness, valence, and babiness). None of these are captured in the model.

Another factor to be considered is the role of intrinsic (innate) differences in children's information processing abilities. In this chapter we have presented a model that explains inter-individual variation in processing speed and working memory capacity as the consequence, not the cause, of differences in vocabulary size. However, there are a number of reasons to believe that there are also intrinsic or very early developing differences in infant's information processing skills, that will themselves have a causal effect on the speed of vocabulary acquisition. For example, Chonchaiya, Tardif, Mai, Xu, Li, Kaciroti, Kileny et al. (2013) showed that improvement in rapid auditory processing capabilities at the subcortical level (in response to masked clicks) between 6 weeks and 9 months of age predicted vocabulary size at 9 months of age and Çetinçelik et al. (2020) demonstrated that 10-month-old infants' neural tracking of the rhythm of speech at the syllable rate predicted vocabulary size 8 months later. There is also evidence that children who learn to segment words from running speech early are at an advantage in vocabulary acquisition (Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014). We also know that there is a strong genetic component to individual differences, with behavioural genetics studies suggesting that a significant proportion of the relation between parental input and child language can be attributed to shared genetic effects (Dale, Tosto, Hayiou-Thomas, & Plomin, 2015). Thus, the most plausible picture, to our minds, is one in which small intrinsic differences in processing capacity are then either magnified or minimised by differences in the child's environment throughout development. Children with slower processing speeds, smaller processing windows, or slower rates of learning at birth may be at an initial disadvantage, but these differences are amplified by impoverished input, or mitigated by enriched input, throughout the early years.

5. Conclusion

A computational model that implements the principles of usage-based theories can simulate inter-individual variation in English vocabulary acquisition. The results suggest that exposure to linguistic input, read through a fixed capacity processing constraint, leads associative learners to store phonological knowledge in chunks of information of varying size, and this stored knowledge influences both non-word repetition performance, processing speed, and vocabulary growth. However, more work is needed to investigate the role of other factors such as semantic knowledge, and to explore the role that early learned or intrinsic differences in domain-general processing capacities play in explaining the speed of vocabulary growth in the first three years of life.

Postscript: Tribute to Heike Behrens from the first author

Heike's work had a huge influence on me when I was a PhD student and she was a postdoctoral researcher, then at Mike Tomasello and Elena Lieven's Max Planck department in Leipzig, back in the 1990s and 2000s. This was an exciting time in which the usage-based theory was really taking off, and Heike's work on grammar acquisition – especially that with Elena Lieven and Kirsten Abbot-Smith – played a huge role in developing my own ideas about how children learn language, as well as critically forcing all of us to think about other languages beyond English. Since then she has become a valuable colleague, as a guest researcher at our Max Planck Institute in Nijmegen, and a firm friend. The focus of this chapter – individual differences in vocabulary acquisition – is not something that Heike has worked on extensively. But it is my firm contention that individual differences can only be properly understood within a constructivist approach of the kind proposed by Heike. This chapter is inspired by her work.

Acknowledgments

We would like to thank all of the families who participated in the Language 0–5 Project. This work was supported by the ESRC International Centre for Language and Communicative Development (LuCiD), funded by the UK Economic and Social Research Council [ES/L008955/1].

References

doi Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking quality and quantity of parental linguistic input to child language skills: A meta-analysis. *Child Development*, 92(2), 484–501.

doi Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics* 47(2), 383–411.

Behrens, H. (2017). The role of analogy in language acquisition. In M. Hundt, S. Mollin, & S. Pfenniger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 215–239). Cambridge University Press.

Behrens, H. (2021). Constructivist approaches to first language acquisition. *Journal of Child Language*, 48(5), 959–983.

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C.F., Ramírez-Esparza, N., R.Hamrick, L., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., Alphen, P. van, & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, 120(52), e2300671120.

Borovsky, A. (2022). Lexico-semantic structure in vocabulary and its links to lexical processing in toddlerhood and language outcomes at age three. *Developmental Psychology*, 58(4), 606–630.

Braginsky, M., Yurovsky, D., Marchman, V.A., & Frank, M.C. (2019). Consistency and variability in children's word learning across languages. *Open Mind (Camb)*, 3, 52–67.

Case, R., Kurland, D.M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33(3), 386–404.

Çetinçelik, M., Rowland, C.F., & Snijders, T.M. (2020). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology*, 11, 589096.

Chonchaiya, W., Tardif, T., Mai, X., Xu, L., Li, M., Kaciroti, N., Kileny, P.R., Shao, J., & Lozoff, B. (2013). Developmental trends in auditory processing can provide early predictions of language acquisition in young infants. *Developmental Science*, 16(2), 159–172.

Cowan, N. (2022). Working memory development: A 50-year assessment of research and underlying theories. *Cognition*, 224, 105075.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, 85(4), 1330–1345.

Dale, P.S., Tosto, M.G., Hayiou-Thomas, M.E., & Plomin, R. (2015). Why does parental language input style predict child language development? A twin study of gene-environment correlation. *Journal of Communication Disorders*, 57, 106–117.

Ellis, N. (2026). What's in a word. In K. Madlener-Charpentier, M. Verspoor, M. Weder, & A. Häcki Buhofer (Eds.), *Usage-based perspectives on language and language acquisition. In honour of Heike Behrens*. John Benjamins (this volume)

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.). Paul H Brookes.

Fernald, A., & Marchman, V.A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, 83(1), 203–222.

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

Frank, M.C., Braginsky, M., Yurovsky, D., & Marchman, V.A. (2021). *Variability and consistency in early language learning: The Wordbank project*. The MIT Press. https://langcog.github.io/wordbank-book/.

Frank, M.C., Braginsky, M., Yurovsky, D., & Marchman, V.A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694.

Freudenthal, D., Gobet, F., & Pine, J.M. (2024). MOSAIC+: A crosslinguistic model of verb-marking errors in typically developing children and children with developmental language disorder. *Language Learning*, 74(1), 111–145.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5), 752–775.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17, 684–691.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four … Or is it two? *Memory*, 12(6), 732–747.

Gobet, F., Lane, P., Croker, S., Cheng, P., Jones, G., Oliver, I., & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.

Gobet, F., & Simon, H.A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, 6(3), 225–255.

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, Language, and Hearing Research*, 50(1), 177–195.

Hasson, U., Chen, J., & Honey, C.J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19(6), 304–313.

Hurtado, N., Marchman, V.A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, 33, 227–249.

Jessop, A., Pine, J., & Gobet, F. (2025). Chunk-based incremental processing and learning: An integrated theory of word discovery, implicit statistical learning, and speed of lexical processing. *Psychological Review*, 132(6), 1340–1374.

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.

Jones, G. (2012). Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in Psychology*, 3, 167.

Jones, G. (2011). A computational simulation of children's performance across three nonword repetition tests. *Cognitive Systems Research*, 12(2), 113–121.

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(2), 298–310.

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(6), 853–873.

Jones, G., & Rowland, C. F. (2017). 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. *Cognitive Psychology*, 98, 1–21.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, 22(2), 154–169.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 802–823.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*, 52(6), 1401–1416.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, 8, 1164.

Marchman, V. A., Ashland, M. D., Loi, E. C., Munevar, M., Shannon, K. A., Fernald, A., & Feldman, H. M. (2023). Associations between early efficiency in language processing and language and cognitive outcomes in children born full term and preterm: Similarities and differences. *Child Neuropsychology*, 29(6), 886–905.

Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish – English bilinguals. *Journal of Child Language*, 37(4), 817–840.

Marchman, V. A., Loi, E. C., Adams, K. A., Ashland, M., Fernald, A., & Feldman, H. M. (2018). Speed of language comprehension at 18 months old predicts school-relevant outcomes at 54 months old in children born preterm. *Journal of Developmental and Behavioral Pediatrics: JDBP*, 39(3), 246–253.

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, 22(3), 491–495.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33, 1619–1624.

Meints, K., Fletcher, K., & Just, J. (2017). *The Lincoln Toddler Communicative Development Inventory: A UK adaptation of the MacArthur-Bates Communicative Development Inventory: Words and sentences*. University of Lincoln.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352.

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, 11(3), 520–535.

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, 2, 345.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology*, 115, 101238.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rowland, C. F., Bidgood, A., Jones, G., Jessop, A., Stinson, P., Pine, J. M., Durrant, S., & Peter, M. S. (2025). Simulating the relationship between nonword repetition performance and vocabulary growth in 2-year-olds: Evidence from the Language 0–5 project. *Language Learning*, 75(2), 379–423.

Rowland, C.F., Bidgood, A., Durrant, S., Peter, M., & Pine, J. M. (2018). *The Language 0-5 Project. University of Liverpool.* Available from https://osf.io/kau5f/.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research: JSLHR*, 47(1), 223–234.

Simon, H.A. (1974). How big is a chunk? *Science*, 183(4124), 482–488.

Skoruppa, K. (2026). Usage and context in pediatric speech and language therapy: Helping children with oral developmental language disorder build their language system(s). In K. Madlener-Charpentier, M. Verspoor, M. Weder, & A. Häcki Buhofer (Eds.), *Usage-based perspectives on language and language acquisition. In honour of Heike Behrens*. John Benjamins (this volume).

Slobin, D. (2026). The wealth of the stimulus. In K. Madlener-Charpentier, M. Verspoor, M. Weder, & A. Häcki Buhofer (Eds.), *Usage-based perspectives on language and language acquisition. In honour of Heike Behrens*. John Benjamins (this volume).

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, 12(3), 369–373.

Ståhlberg-Forsén, E., Latva, R., Leppänen, J., Lehtonen, L., & Stolt, S. (2022). Eye tracking based assessment of lexical processing and early lexical development in very preterm children. *Early Human Development*, 170, 105603.

Stärk, K., Kidd, E., & Frost, R. L. A. (2022). The effect of children's prior knowledge and language abilities on their statistical learning. *Applied Psycholinguistics*, 43(5), 1045–1071.

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23–36.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory & Cognition*, 13(3), 193–201.