

# Psychological Review

## **Chunk-Based Incremental Processing and Learning: An Integrated Theory of Word Discovery, Implicit Statistical Learning, and Speed of Lexical Processing**

Andrew Jessop, Julian Pine, and Fernand Gobet

Online First Publication, May 12, 2025. <https://dx.doi.org/10.1037/rev0000564>

### CITATION

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*. Advance online publication. <https://dx.doi.org/10.1037/rev0000564>

# Chunk-Based Incremental Processing and Learning: An Integrated Theory of Word Discovery, Implicit Statistical Learning, and Speed of Lexical Processing

Andrew Jessop<sup>1, 2</sup>, Julian Pine<sup>1, 2</sup>, and Fernand Gobet<sup>2, 3, 4</sup>

<sup>1</sup> Department of Psychology, University of Liverpool

<sup>2</sup> Economic and Social Research Council International Centre for Language and Communicative Development

<sup>3</sup> Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science

<sup>4</sup> School of Psychology, University of Roehampton

According to chunking theories, children discover their first words by extracting subsequences embedded in their continuous input. However, the mechanisms proposed in these accounts are often incompatible with data from other areas of language development. We present a new theory to connect the chunking accounts of word discovery with the broader developmental literature. We argue that (a) children build a diverse collection of chunks, including words, multiword phrases, and sublexical units; (b) these chunks have different processing times determined by how often each chunk is used to recode the input; and (c) these processing times interact with short-term memory limitations and incremental processing to constrain learning. We implemented this theory as a computational modeling architecture called Chunk-Based Incremental Processing and Learning (CIPAL). Across nine studies, we demonstrate that CIPAL can model word discovery in different contexts. First, we trained the model with 70 child-directed speech corpora from 15 languages. CIPAL gradually discovered words in each language, with cross-linguistic variation in performance. The model's average processing time also improved with experience, resembling the developmental changes observed in children's speed of processing. Second, we showed that CIPAL could simulate seven influential effects reported in statistical learning experiments with artificial languages. This included a preference for words over nonwords, part words, frequency-matched part words, phantom words, and sublexical units. On this basis, we argue that incremental chunking is an effective implicit statistical learning mechanism that may be central to children's vocabulary development.


**Keywords:** chunking, word discovery, speed of processing, statistical learning, computational modeling


**Supplemental materials:** <https://doi.org/10.1037/rev0000564.supp>


When listening to a conversation in an unfamiliar language, we find it difficult to pinpoint the individual words as there are no reliable acoustic boundaries that separate them in fluent speech. Infants learning their first language face a similar problem. The speech they hear is mainly produced in concatenated bursts of multiple words without any pauses (Cole & Jakimik, 1980; Junge, 2017). Yet, regardless of the specific languages they are learning, infants still

manage to build a lexicon and gradually become productive language users. By the time they reach their first birthday, most children have already acquired a small vocabulary of high-frequency content words that they can recognize in different contexts (Bergelson & Swingley, 2012, 2015, 2018; Lany et al., 2018). This is followed by a rapid growth in their expressive vocabulary (e.g., McMurray, 2007) and their ability to recognize familiar words (e.g., Fernald et al., 1998)

Hongjing Lu served as action editor.

Andrew Jessop  <https://orcid.org/0000-0002-2207-4663>

Julian Pine  <https://orcid.org/0000-0002-7077-9713>

Fernand Gobet  <https://orcid.org/0000-0002-9317-6886>

Additional online material relating to this research are available in the Open Science Framework repository at <https://osf.io/fhrxg/>. This research was funded by the Economic and Social Research Council (Grants ES/L008955/1 and ES/S007113/1) as part of the International Centre for Language and Communicative Development (<https://www.lucid.ac.uk/>).

The authors are grateful to the following researchers for providing official adaptations of the Communicative Development Inventory: Caroline Rowland (Dutch; English U.K.), Tiia Tulviste (Estonian), Ciara O'Toole (Irish), and Darinka Anđelković (Serbian).

Open Access funding provided by The University of Liverpool: This work

is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0>). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Andrew Jessop played a lead role in conceptualization, formal analysis, investigation, methodology, software, visualization, and writing—original draft. Julian Pine played a supporting role in conceptualization and methodology and an equal role in funding acquisition, supervision, and writing—review and editing. Fernand Gobet played a supporting role in conceptualization and methodology and an equal role in funding acquisition, supervision, and writing—review and editing.

Correspondence concerning this article should be addressed to Andrew Jessop, Department of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, United Kingdom. Email: [andrew.jessop@liverpool.ac.uk](mailto:andrew.jessop@liverpool.ac.uk)

throughout their second year. This means that infants can discover words in their language input, even though most of the words they hear are embedded in longer utterances. But how do they do this?

We propose that children use an associative learning mechanism called chunking (Gobet et al., 2001), where sequences of linguistic elements are grouped into units (e.g., *b, i, d, a, k, u* → *bidaku*). Unlike previous accounts that limit the number of chunks stored in long-term memory (LTM; e.g., Brent & Cartwright, 1996; Perruchet & Vinter, 1998), we argue that children build a diverse collection of chunks, including words, multiword phrases, and sublexical units. We also suggest that these chunks have different representational strengths that determine their processing costs, allowing regularly accessed chunks to have faster retrieval times, resulting in an interaction between experience and memory constraints. We implement this theory as a computational architecture called Chunk-Based Incremental Processing and Learning (CIPAL) and model word discovery in 70 child-directed speech corpora from 15 different languages. We then use this architecture to simulate the behavioral patterns observed in statistical learning experiments with artificial languages. Before describing these studies in detail, we first review evidence for chunking as an implicit statistical learning mechanism and how it could overcome the challenges of word discovery.

## Statistical Learning and Word Discovery

There are often multiple ways utterances can be segmented. For example, the sentence “hi doggie” could be represented with six phonemes (/h/, /a/, /d/, /ɒ/, /g/, /l/), three syllables (/haɪ/, /dɒ/, /gɪl/), two words (/haɪ/, /dɒgɪl/), or one multiword unit (/haɪdɒgɪl/). It could also be parsed into units that cross word boundaries, such as (/h/, /aɪd/, /ɒgɪl/). Since pauses tend to occur at the boundaries of utterances rather than words, infants need to exploit other sources of information to find the words among these other possible partitions. Although many potential cues have been identified (e.g., lexical stress; Thiessen & Saffran, 2003, 2007), analyses have found that the most informative features for predicting word boundaries vary between languages (Jarosz & Johnson, 2013). Since infants do not know in advance which specific cues are the most informative, they must discover their first words using features that can be perceived and exploited in any language without prior knowledge. One option is the distributional structures that guide how linguistic elements are organized (Harris, 1954, 1955).

Across different languages, syllable pairs that appear together in the same word tend to have stronger statistical relationships than pairs that cross word boundaries (Harris, 1955; Saksida et al., 2017). There is clear evidence that adults and preverbal infants can recognize word-like units after listening to continuous input, where statistical properties are the only differentiating feature. In their seminal study, Saffran, Aslin, and Newport (1996) trained 8-month-olds with artificial languages containing four trisyllabic “words” (e.g., *golatu*, *daropi*) that were repeated in a random order for 2 min. This exposure was produced as a continuous stream with no pauses or prosodic cues to mark the boundaries between the individual words. After listening to the language, Saffran, Aslin, and Newport assessed whether the infants could discriminate the words of the language from foils (e.g., *tudaro*, *pigola*) built from syllable pairs with weaker statistical relationships than the words. The strength of these relationships was quantified using transitional probabilities (TPs), which represent the probability of two syllables (e.g., *tu*, *da*) occurring together in the

input. In the forward direction, TPs are calculated by dividing the frequency of the pair (e.g., *tuda*) by the frequency of the first syllable alone (e.g., *tu*). This means that if *tuda* occurs five times, and *tu* is presented 10 times, then the forward TP (FTP) of *tuda* is 0.5. To calculate a backward TP (BTP), the frequency of the pair (e.g., *tuda*) is divided by the frequency of the second syllable (e.g., *da*). In both cases, TPs quantify the consistency of the pairing; FTPs represent the probability that *tu* is followed by *da*; BTPs represent the probability that *da* is preceded by *tu*. In Saffran, Aslin, and Newport’s study, the internal TPs of the test items in each direction were identical. They found that the infants looked for longer when presented with the low-TP foils (0 or 0.33) compared to the high-TP words (always 1.0), demonstrating that they could distinguish items that differed in statistical structure. This ability is called *statistical learning*. It has been observed in multiple replications of Saffran, Aslin, and Newport’s work (see Black & Bergmann, 2017) and in other studies testing different age groups, cognitive modalities, languages, and populations (see Frost et al., 2019; Saffran & Kirkham, 2018). This implies that the statistical structure of natural languages could provide enough information for infants to jumpstart their vocabulary development. Once they have acquired some words, children may identify and exploit language-specific cues, such as the dominant stress patterns or phonotactic constraints on word forms (Mattys et al., 2005).

While there is consensus that statistical regularities in natural languages can help infants discover their first words (Aslin, 2017; Perruchet, 2019; Saffran & Kirkham, 2018), there is debate surrounding the nature of the learning mechanisms involved in this process (Christiansen, 2019; Endress et al., 2020; Perruchet, 2019; Perruchet & Pacton, 2006; Thiessen, 2017). One hypothesis is that infants unconsciously track the statistical patterns in their input to generate probabilistic evidence of where the boundaries between the words are located (Aslin, 2017; Endress & Johnson, 2021; Kuhl, 2004; Saffran & Kirkham, 2018; Swingley, 2005). The central claim of this *statistics-based* theory is that words are initially represented as probabilistic links between individual elements rather than as complete units. High-probability sequences are assumed to be part of the same word (e.g.,  $A_1A_2$ ,  $B_1B_2$ ), whereas transitions high in entropy (e.g.,  $A_2B_1$ ) are interpreted as breakpoints in the speech stream. These representations may then be used to build psychological units, like words, in LTM, but this is regarded as a separate step that comes after learning about the statistical regularities in the input (McCauley & Christiansen, 2019; Perruchet & Pacton, 2006; Saffran, 2001; Slone & Johnson, 2018).

The statistics-based theories predominantly emerged from statistical learning experiments, which routinely use the TPs between adjacent syllables to assess participants’ sensitivity to different input languages and test conditions (e.g., Aslin et al., 1998; Pelucchi et al., 2009a; Saffran, Aslin, & Newport, 1996). Since participants can discriminate items defined by these statistical properties, the statistics-based accounts argue that they must be extracting and building implicit representations of the same statistics (e.g., TPs between syllables). These accounts also argue that this probabilistic knowledge provides infants with information that is both necessary and sufficient to predict the location of words (Aslin, 2017; Aslin et al., 1998; Endress et al., 2020; Saffran, Aslin, & Newport, 1996). For instance, Saffran, Aslin, and Newport (1996) suggested that children learning English can find word boundaries in sequences like *prettybaby* by tracking the TPs between syllables, since the transition

from *pre* to *ty* has a higher probability than from *ty* to *ba*. This idea is compatible with the mechanisms used in simple recurrent network models (Elman, 1990), which can be trained to predict the location of word boundaries by aligning with the TPs of the input (Cairns et al., 1997; Christiansen et al., 1998; Mirman et al., 2010). Corpus analyses have also found that pairwise statistics can accurately predict word boundaries in transcripts of child-directed speech, although different calculation methods are informative in different languages (Gervain & Guevara Erra, 2012; Jarosz & Johnson, 2013; Saksida et al., 2017); for instance, BTPs are the most effective algorithms for finding word boundaries in Hungarian and Polish, whereas FTPs are more effective in Italian.

However, there are several problems with the statistics-based approach to word discovery. First, the statistical cues that correlate with word (or morpheme) boundaries vary between different languages, but it is not clear how the learning mechanism determines which statistics need to be calculated (e.g., FTPs or BTPs), which primitives or linguistic features should be used in these calculations (e.g., syllables or phonemes), and how fluctuations in these statistics should be used to locate words (e.g., absolute/global or relative/local thresholds). Each of these decisions strongly influences the effectiveness of statistics-based word discovery across languages (Gervain & Guevara Erra, 2012; Jarosz & Johnson, 2013; Saksida et al., 2017). Likewise, experiments with artificial languages have demonstrated that infants can discriminate words from foils that have different TPs in only one direction (e.g., backward: Pelucchi et al., 2009a; forward: Pelucchi et al., 2009b), suggesting that they do not rely on one specific statistical property. To explain these data, statistics-based learning mechanisms would need to be flexible and capable of targeting the most informative properties in any given input (Endress et al., 2020; Saffran & Kirkham, 2018), but it is not clear how they would make this determination without feedback or prior knowledge of the language.

The results of multiple experiments have also found that participants' behavior does not always align with statistical variables, which conflicts with the core predictions of statistics-based theories. For instance, studies using visual and linguistic stimuli have found that infants and adults show larger discrimination effects for words (e.g., *daku*) than sublexical items embedded within words (e.g., *gola* from *golabu*), even though these sequences have identical TPs and frequencies, so statistics-based theories would not predict a difference (Giroux & Rey, 2009; Slone & Johnson, 2018). Participants have also shown stronger preferences for words than "phantom" (or "illusory") sequences that have identical TPs to the words but never appeared in the exposure language (Ordin, Polyanskaya, & Soto, 2020; Ordin, Polyanskaya, Soto, & Molinaro, 2020; Perruchet & Poulin-Charronnat, 2012; Polyanskaya, 2022; Slone & Johnson, 2015, 2018; cf. Endress & Langus, 2017; Endress & Mehler, 2009). Instead, the experimental literature suggests that participants incrementally build coherent representations of the sequences they encounter in their input, consistent with the predictions of chunk-based (or memory-based) statistical learning theories (e.g., Perruchet, 2019; Perruchet & Vinter, 1998; Thiessen, 2017).

## Chunking Theories of Word Discovery

Chunk-based theories are a broad category that includes many different accounts of word discovery (see Perruchet, 2019). A general definition of a *chunk* is "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001, p. 236). The idea was first

introduced into memory research by G. A. Miller (1956) to explain how participants can maximize the amount of information in short-term memory (STM) by packing information together (see Norris & Kalm, 2021). The fundamental claims of these theories differ from the statistics-based approaches in two ways. First, chunk-based theorists argue that infants discover words by extracting subsequences that appear in their continuous input, which are stored as atomic units in LTM rather than as correlational links between elements. For example, after hearing a sequence such as *bidakutupir-ogolabubidakupadoti*, infants might extract *bidaku* and represent it as a unified symbol in LTM. They could then use this chunk to partition the input when they reencounter this sequence again: *bidaku tupir-ogolabu bidaku padoti*. This concept of chunked representations is consistent with the findings of several statistical learning studies (see Perruchet, 2019; cf. Endress et al., 2020), as well as evidence from the broader psychology literature suggesting that people learn and utilize chunks across different cognitive domains (Gobet et al., 2001).

The second core argument of chunk-based theories is that children do not perform statistical computations but use other strategies to find the chunks that correspond to words among the plethora of other subsequences in their language input.<sup>1</sup> One hypothesis is that chunking is guided by simplicity principles, where new chunks are created only if they reduce the combined cost of processing the input and storing the chunks in LTM (e.g., Brent & Cartwright, 1996; Goldwater et al., 2009; Robinet et al., 2011). Another is that infants start by building chunks of entire utterances and then use this knowledge to decompose future input into words, working on the assumption that short utterances, including single-word utterances, will often appear as subsequences in longer ones (e.g., Monaghan & Christiansen, 2010). Other accounts have suggested that infants build a variety of chunks of different sizes, which are then strengthened with experience or pruned from memory if they are not regularly used (e.g., Alhama & Zuidema, 2017; Perruchet & Vinter, 1998). Thus, chunking approaches argue that our ability to distinguish sequences with different statistical properties is a consequence of learning and not the mechanism that drives it.

## Problems With Previous Chunking Models

Computational modeling studies have provided concrete demonstrations of how infants could use chunking to find words in their input without relying on statistical computations. Specifically, chunking has been used to accurately segment naturalistic transcripts of parental speech (e.g., Brent & Cartwright, 1996; French et al., 2011; Monaghan & Christiansen, 2010) and replicate the behavioral patterns observed in experiments with artificial languages (e.g., Alhama & Zuidema, 2017; French et al., 2011; Perruchet & Vinter, 1998; Robinet et al., 2011). It has also been used to explain effects reported in the implicit learning literature, including how participants find

<sup>1</sup> Some theories argue that children use statistics to find meaningful chunks in the language (e.g., the Chunk-Based Learner model uses BTPs; McCauley & Christiansen, 2019). We argue that these models should be categorized as statistics-based theories rather than chunking theories in discussions of word discovery since their predictions are consistent with the former. For instance, these accounts suggest that participants would not distinguish words from phantom or sublexical units since they rely on pair-wise statistics to select chunks. Therefore, we use the term "chunking theories" as an abbreviation for "chunking-without-statistics theories." A similar distinction was made by Perruchet (2019), who described such models as hybrid theories.



rules in artificial grammars (e.g., Servan-Schreiber & Anderson, 1990), why chess grandmasters can recall board configurations more accurately than less experienced players (e.g., Gobet, 1998), why children make certain morphosyntactic errors during development (e.g., Freudenthal et al., 2007), and why children find it easier to recall nonword sequences when they resemble real words in their language (e.g., Jones et al., 2007).

Although these data provide clear evidence to support chunk-based approaches, the simulations in these studies were conducted with several distinct modeling frameworks that use different strategies to extract chunks from the input (e.g., Retention and Recognition: Alhama & Zuidema, 2017; Truncated Recursive Autoassociative Chunk Extractor: French et al., 2011; Phonotactics from Utterances Determine Distributional Lexical Elements [PUDDLE]: Monaghan & Christiansen, 2010; PARSE: Perruchet & Vinter, 1998; Minimum Description Length Chunker: Robinet et al., 2011). These models often make fundamentally different assumptions and predictions about the way that infants learn and use chunks, which have consequences when viewed in a broader developmental context.

For example, the PUDDLE architecture (Monaghan & Christiansen, 2010) assumes that children's language input is presented in discrete utterances: A small group of words are produced together in a short continuous burst, followed by a pause that unambiguously marks the boundary between different groups of words (i.e., "puddles" of sound). It also assumes that many of these utterances will contain only one word (e.g., "where," "kitty"). On this basis, PUDDLE uses a coarse-to-fine chunking strategy; the model initially builds chunks for entire utterances and uses this knowledge as a stepping stone for learning the individual words (e.g., "lookkitty" → "look," "kitty"). As the model's knowledge grows, it will start to recognize familiar chunks in the input—particularly those extracted from single-word utterances—and will use these chunks to parse multiword sequences and discover new words. For example, once PUDDLE has built chunks for "where" and "kitty," it will use these chunks to segment "where is kitty hiding" and learn the unfamiliar parts as new chunks (i.e., "is" and "hiding"). PUDDLE also exploits utterance boundaries to extract reliable information about the legal phonotactics of the language, which are used as an additional constraint on word segmentation. Specifically, the model can only use a chunk to parse the input when the sequence is flanked by diphones the model has previously encountered at the beginning or end of words.

Simulations with PUDDLE have found that it can discover words in transcripts of naturalistic child-directed speech from many different languages (Caines et al., 2019; Monaghan & Christiansen, 2010). It also appears to be more effective than statistics-based approaches that use TPs or mutual information to find word boundaries (Cabiddu et al., 2023; Caines et al., 2019). However, since the model assumes that the input is structured as a series of discrete utterances, it cannot explain how participants learn to discriminate words from foil sequences in experiments with artificial languages (e.g., Aslin et al., 1998; Saffran, Aslin, & Newport, 1996). In the canonical statistical learning paradigm, participants listen to a stream of syllables for several minutes without any pauses (i.e., a "sea" of sound). When the language is presented as one long utterance, PUDDLE cannot extract the individual words or other subsequences. Instead, the model stores the entire exposure as one chunk.<sup>2</sup>

In contrast, Perruchet and Vinter (1998) developed the PARSE architecture to demonstrate that the preferences observed in statistical learning studies (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) could be explained using chunking. The

model assumes that repetitive subsequences in the input are more likely to be words than foils. It begins by randomly segmenting the material into groups of 1–3 units (e.g., *ba bu pu bu pa da du ta ba* → *ba bupubu pada du taba*). It then checks whether these random groupings (e.g., *taba*) match a chunk in its memory, called the *perception shaper*, and creates a new chunk if there is no match. Since PARSE learns new chunks by randomly clustering the material it encounters, it needs a way to identify which chunks correspond to words. To do this, each chunk has a weight that quantifies its representational strength. When a chunk is used to segment the input, it is reinforced and becomes stronger, while the other chunks that are not in use become weaker through decay. The unused chunks also become weaker through interference, receiving an additional penalty when they contain any of the elements or subsequences that also appear in the selected chunks (e.g., *ta* and *ba* become weaker when *taba* is selected). Ultimately, chunks are removed from the perception shaper if they are not reinforced before their weight drops to zero. In the early stages of learning, many of the chunks in the perception shaper contain sequences that cross word boundaries. However, through reinforcement, decay, interference, and a bias for larger chunks, only repetitive sequences (typically words) are retained by the model with experience.

Several studies have found that PARSE is effective at identifying words in artificial languages that contain a small number of word types repeated at regular intervals in a random order (e.g., Giroux & Rey, 2009; Perruchet & Poulin-Charronnat, 2012; Perruchet & Vinter, 1998). However, many word types in natural languages have a low frequency (Piantadosi, 2014) and contain sublexical patterns that appear in other words (e.g., rhymes, affixes). PARSE will often discard chunks for these patterns before they reappear in the input, since its decay and interference mechanisms retain only the most frequent and distinctive sequences.<sup>3</sup> In general, decay mechanisms may be detrimental to word discovery in natural languages; pilot studies with the PUDDLE architecture found that a decay function resulted in a small lexicon and poor performance with English child-directed speech (Monaghan & Christiansen, 2010).

## A New Chunking Model of Word Discovery

Most computational models of word discovery are calibrated for either large naturalistic corpora (e.g., PUDDLE) or small artificial languages (e.g., PARSE). Only a minority of architectures has demonstrated an ability to find words in both contexts (e.g., Truncated Recursive Autoassociative Chunk Extractor: French et al., 2011). This raises the important question of whether infants are using different mechanisms to discover words in statistical learning experiments and their native language, given that the artificial languages used in these studies do not contain many of the features that appear in natural languages (e.g., utterance boundaries, lexical stress).

To address this issue, Pelucchi et al. (2009a) trained English-learning 8-month-old infants with a set of 12 carefully selected but

<sup>2</sup> We demonstrate this behavior in our additional online material using one of the artificial languages from Saffran, Aslin, and Newport's (1996) second experiment (see *notebooks/puddle-examples.html* at <https://osf.io/fhrxg/>). We also show that PUDDLE will fail to segment multiword sequences, when there are no one-word utterances in the input.

<sup>3</sup> We demonstrate this in our additional online material using a corpus of English child-directed speech (see *notebooks/parser-examples.html* in <https://osf.io/fhrxg/>).

naturalistic Italian utterances, which were produced with the prosody and intonation contours that characterize infant-directed speech. Afterward, the infants heard two words that each occurred six times in the input (e.g., *fuga, melo*) and two novel Italian words that did not appear as complete sequences in the exposure (e.g., *pane, tema*). Pelucchi et al. (2009a) observed a reliable familiarity preference, as the infants showed longer looking times in trials with familiar words than novel words. This was consistent with the word and nonword comparisons in studies with artificial languages (Saffran, Aslin, & Newport, 1996). In another experiment, the same authors found that infants also showed longer looking times for high-TP (1.0) than low-TP words (0.33), even though the words in both conditions occurred 18 times in the training material. This was similar to the novelty preferences reported in Aslin et al.'s (1998) experiment with artificial languages, which showed that infants could discriminate high-TP words from low-TP part words that each appeared 45 times in the language.

From these findings, it seems likely that infants use the same learning strategies to discover words both in their native language and the experimental languages used in statistical learning studies; at least, there is no concrete evidence to suggest otherwise. If this is true, then computational models should be capable of simulating word discovery in both environments. Our aim was to construct a new computational model that can discover words in both child-directed speech and artificial languages using chunking mechanisms that are compatible with the broader developmental literature.

To demonstrate that a model can extract words from child-directed speech, it is important to test it on a diverse set of languages. Previous studies using both statistics-based and chunking algorithms have consistently found cross-linguistic variation in performance (Batchelder, 2002; Caines et al., 2019; Fourtassi et al., 2013; Gervain & Guevara Erra, 2012; Jarosz & Johnson, 2013; Phillips & Pearl, 2014; Saksida et al., 2017). For example, models that can identify English words with over 80% accuracy have shown less than 60% accuracy in other languages (Fourtassi et al., 2013; Saksida et al., 2017). It is possible that word discovery is intrinsically harder in some languages due to differences in average word length, type-token ratio, linguistic rhythm, syllable complexity, or other properties (Caines et al., 2019; Gervain & Guevara Erra, 2012; Saksida et al., 2017). Yet, regardless of the specific languages they are learning, most children gradually build a lexicon and become fluent language users within a few years. Their ability to discover words in continuous speech is robust and can cope with the many ways in which languages vary. For this reason, our first goal was to demonstrate that our new chunking framework could extract words from child-directed speech corpora in several different languages without making any adjustments to the model.

Following Saffran et al.'s initial studies of infants (Saffran, Aslin, & Newport, 1996) and adults (Saffran, Newport, & Aslin, 1996), many experiments have investigated statistical learning using variations of the original paradigm (see Frost et al., 2019; Saffran & Kirkham, 2018). Our second goal was to simulate some of the most reliable and influential findings in this literature, including studies testing the predictions of the statistics-based and chunking accounts. First, we tested whether the model could distinguish words from illegal nonwords, low-frequency part words, and high-frequency part words, using the languages created by Saffran, Newport, and Aslin (1996) and Aslin et al. (1998). These studies were influential in the development of statistics-based accounts, as the results imply that

infants are sensitive to distributional cues in the input. We then tested whether the model could identify words designed to have higher TPs in either the forward or backward direction, which has been observed in studies with infants and adults (Pelucchi et al., 2009a, 2009b; Perruchet & Desauty, 2008). Finally, we tested whether our new model showed a preference for words over phantom units (Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012; Slone & Johnson, 2018) and sublexical units (Giroux & Rey, 2009; Slone & Johnson, 2018). The statistics-based and chunk-based theories make opposing predictions for these studies, so they are central to discussions of the nature of statistical learning.

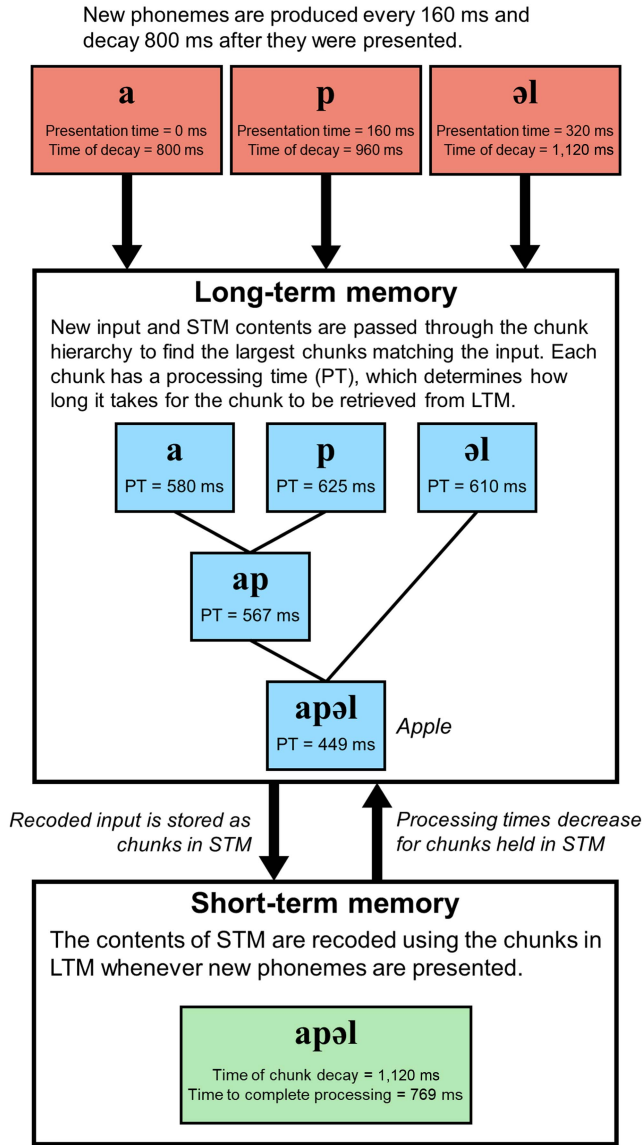
Building on previous work, we developed the CIPAL architecture. CIPAL is different to other chunking models of word discovery (e.g., *PARSER*, *PUDDLE*) in three important ways. First, the model incrementally and concurrently accumulates a large quantity of chunks in a hierarchical LTM, starting with the phonemes of the language before progressing to sublexical, lexical, and multiword units. Second, each chunk has a dynamic strength of representation, which is implemented as a processing time that gets faster whenever the chunks are used to recode the input. These timings interact with the model's incremental processing and limited STM capacity to constrain learning. Finally, the model is designed to be cognitively plausible and compatible with the broader developmental literature: CIPAL has a limited STM capacity (e.g., Cowan, 2001), it processes language incrementally (e.g., Tanenhaus et al., 1995), it learns different types of chunks (e.g., Jones et al., 2007), and these chunks get faster with experience (e.g., Fernald et al., 1998).

### CIPAL: Chunk-Based Incremental Processing and Learning

CIPAL is an integrated theory of word discovery, implicit statistical learning, and speed of lexical processing, implemented as a cognitive architecture for building process models (Jarecki et al., 2020). It is based on the EPAM/CHREST framework (Elementary Perceiver and Memoriser/Chunk Hierarchy and RETrieval STRuctures; de Groot & Gobet, 1996; Feigenbaum & Simon, 1984) and the CLASSIC model of vocabulary development (Chunking Lexical and Sub-lexical Sequences in Children; Jones et al., 2005, 2007, 2014), with similarities to the *PARSER* model of statistical learning (Perruchet & Vinter, 1998). The CIPAL architecture has a STM that temporarily holds the active stimuli, and a LTM that stores familiar patterns as chunks. These memory structures work together to process the input and learn new representations. As new material is presented, it is recoded using the largest available chunks in LTM (e.g., *d, a, k, u* → *da, ku*) and then stored in STM. If two or more chunks are needed to represent the input, CIPAL attempts to cluster adjacent units together and add them to LTM as a new chunk (e.g., *da, ku* → *daku*). This creates a cyclical relationship between processing and learning, where chunks are used to compress the input stored in STM, and then build new representations in LTM. Figure 1 provides an overview of how a trained model would process the word "apple" by passing the language material between LTM and STM.

The CIPAL theory makes three assumptions about word discovery: (a) When listening to a (natural or artificial) language, we build chunks of different sizes to represent the patterns in the input at multiple levels; (b) each chunk has a representational strength (i.e., a weight or activation level) determined by the number of times it has

**Figure 1**  
*The Main Components of the CIPAL Architecture*



*Note.* The example model in this figure was trained with the word “apple” 40 times (see [notebooks/cipal-examples.html](https://notebooks.cipal-examples.html) in the additional online material at <https://osf.io/fhrxg/>). LTM = long-term memory; STM = short-term memory; CIPAL = Chunk-Based Incremental Processing and Learning. See the online article for the color version of this figure.

been used to process the input, which manifests as a reaction time that can be measured in laboratory experiments; (c) chunk learning is constrained by the interaction of incremental language processing, rapidly decaying memory traces, and chunk-specific reaction times. In the following sections, we explain the developmental evidence for each of these assumptions and describe how they are implemented as cognitive processes in the CIPAL architecture. We then provide a detailed algorithmic description of the model and examples of how these processes allow CIPAL to discover words (and other sequences) in continuous language input. Finally, we explain the

parameter settings that we have used throughout the simulations in this work.

### Assumption 1: We Learn the Patterns of the Language by Building a Diverse Chunk Hierarchy

Many chunking models of word discovery assume that infants have an economy of representation and only learn the chunks that will help them to segment their input (e.g., Brent & Cartwright, 1996; Goldwater et al., 2009; Perruchet & Vinter, 1998). For instance, PARSER only preserves repetitive sequences that are regularly accessed, while neglected chunks are discarded through decay and interference (Perruchet & Vinter, 1998). These algorithms are effective at modeling word segmentation since chunks for illegal sequences or meaningful nonlexical units are either avoided or eventually forgotten. Consequently, they also predict that most of the chunks that children retain over time will be words. However, children appear to build chunks for at least two other aspects of language.

First, several studies have observed that participants are sensitive to the frequency of multiword sequences, even when controlling for word-level frequencies (Armon & Snider, 2010; Bannard & Matthews, 2008; Real & Christiansen, 2007; Tremblay et al., 2011). For example, Bannard and Matthews (2008) tested whether 2- and 3-year-olds could repeat high (e.g., *a drink of milk*) and low frequency (e.g., *a drink of tea*) multiword sequences, where the final words (*milk*, *tea*) and bigrams (*of milk*, *of tea*) occurred with similar frequencies in child-directed speech. They found that the children in both age groups were faster and more accurate at repeating the high-frequency combinations, suggesting that they were sensitive to the familiarity of the complete utterances and were not just processing the individual words. This implies that children learn common multiword sequences as chunks, which show similar frequency effects to other aspects of language (Ambridge et al., 2015; Brysbaert et al., 2018).

Second, there is evidence that children learn chunks for sublexical patterns. Mintz (2013) familiarized English-learning infants with novel words that ended with either the English suffix *-ing* (e.g., *lerjoving*) or a pseudo-affix (e.g., *-ot*, *-dut*). The infants then listened to the novel stems without the suffix (e.g., *lerjov*) as their looking times were measured with the head-turn preference procedure. Mintz found that 15-month-olds could distinguish items that appeared with *-ing* from those that occurred with the novel suffix, although 8-month-olds showed no reliable preference. This suggests that children build chunks for sublexical morphemes that allow them to segment the unfamiliar stem from the familiar suffix (see also Dahan & Brent, 1999). Other evidence for the importance of sublexical chunks comes from studies with nonword repetition tasks, where children are presented with nonsense words and try to repeat them accurately. These studies have found higher repetition accuracy when the nonwords are built from high-frequency syllable patterns (Gathercole, 1995; Jones et al., 2007), which suggests that children retain sublexical chunks that help them to process and learn from unfamiliar sequences.

On this basis, the CIPAL theory assumes that people continually expand their knowledge with experience and ultimately build a diverse collection of chunks of different sizes. Through contact with the language, the model gradually constructs a chunk hierarchy in LTM, where information is represented at multiple levels. For example, if the model were repeatedly presented with *bidaku* as a sequence of syllables, it would first learn the primitives of the sequence (e.g., *bi*, *da*,



*ku*) before building progressively larger chunks (e.g., *bi* → *bida* → *bidaku*) until the entire pattern is represented as a single unit. It does this by concatenating two adjacent units stored in STM into one larger chunk (e.g., *bi* + *da* → *bida*; *bida* + *ku* → *bidaku*). Unlike PARSER, CIPAL does not remove any chunks from LTM. Instead, by iteratively combining two individual chunks into a single unit, the model can recognize patterns of different sizes in the input and simultaneously store chunks for lexical (e.g., *walked*), sublexical (e.g., *-ed*), and multiword units (e.g., *theywalked*). As well as making the model developmentally plausible, this hierarchy of chunks allows CIPAL to efficiently recognize patterns as the utterance unfolds. If unknown sequences are encountered, the model attempts to learn new chunks so that it can process these patterns more efficiently (i.e., with fewer chunks) if they reappear in the input.

### Assumption 2: Each Chunk Has a Dynamic Processing Speed

During their second year, children show a rapid increase in the number of words they can understand and produce (e.g., McMurray, 2007). At the same time, they also become faster and more accurate at recognizing familiar words (e.g., Fernald et al., 1998). This is often assessed using the looking-while-listening (LWL) procedure. In a standard LWL experiment, children are presented with images of two familiar objects on-screen (e.g., a ball and a shoe) and they hear child-directed speech that names one of the items (e.g., “Where is the ball?”). The participant’s eye movements are recorded throughout the experiment and analyzed offline. The latency between the onset of the target word (e.g., “ball”) and the child’s first fixation to the corresponding visual image is calculated as their speed of processing for the trial, with the requirements that they were not already fixating on the target image and their gaze shift occurred within a pre-determined window (e.g., 300–1,800 ms).

Many studies have used the LWL task to study the development of children’s word comprehension abilities alongside their vocabulary. There are three consistent findings in this research: First, children get faster and more efficient in responding to familiar nouns as they get older (e.g., Fernald et al., 1998); second, individual differences in speed of processing correlate with vocabulary size and growth rate (e.g., Fernald et al., 2006; Peter et al., 2019); and third, children who receive larger amounts of child-directed speech and parental interaction tend to have faster response times and larger vocabulary sizes over development (e.g., Hurtado et al., 2008; Weisleder & Fernald, 2013). Some longitudinal studies have observed all three effects within a single sample. For instance, Hurtado et al. (2008) recorded 18-month-old Spanish-learning children in their homes during a 20-min interactive play session with their mothers. They measured the number of words the mothers produced during the session, which ranged from 168 to 1,204 tokens. Six months later, they tested the same infants in the LWL task and measured their vocabulary size with a checklist completed by the parents. The children who heard more word tokens during the play sessions at 18 months showed faster reaction times in the LWL task and larger expressive vocabularies at 24 months. Collectively, the speed of processing literature suggests that early language experience shapes children’s vocabulary knowledge and processing skills.

An important feature of the LWL paradigm is that the participants are tested with high-frequency target words that typically appear in their early lexicons (e.g., *doggie*, *baby*, *ball*, and *shoe*). Some studies

have even used parental questionnaires to exclude trials using words the child does not know (e.g., Fernald et al., 2006). From the perspective of chunking theories, this means that the children are likely to have chunks for the items in the task. Yet, these studies have still observed reliable individual differences in children’s reaction times that correlate with their language experience and vocabulary size. This suggests there is meaningful variance in their lexical processing skills beyond the acquisition of chunks. One explanation is that children have a central processing ability that affects their reaction times for all chunks in their LTM (Donnelly & Kidd, 2020). However, studies with bilingual samples have found that children’s processing speed correlates with their level of experience and vocabulary size within each language, with no reliable correlations in their reaction times across different languages (Hurtado et al., 2014; Marchman et al., 2010). For this reason, we argue that LWL latencies are determined by the strength of the individual chunks used to process the target words rather than a global processing capacity. We suggest that children become faster at retrieving specific chunks from LTM when they are used to process the input. As children become more experienced with the language, their overall speed of processing will improve, driven by the average of multiple chunk-specific processing times for different words in the language.

Several chunking theories have considered the impact of chunk strength on learning and processing (e.g., Monaghan & Christiansen, 2010; Perruchet & Vinter, 1998; Servan-Schreiber & Anderson, 1990). For instance, the chunks in PARSER (Perruchet & Vinter, 1998) are assigned an initial weight (1.0), which increases whenever the chunk is used to segment the input (reinforcement: +0.5) but also decreases when it is not in use (decay: −0.05) or when it shares syllables with the active chunks (interference: −0.005). Units are removed from the model’s memory (called the perception shaper) when their weight falls to zero, which means only the most repetitive and distinctive sequences are stored over time. Similarly, each chunk in PUDDLE (Monaghan & Christiansen, 2010) has an activation level. New chunks are assigned an initial activation level of 1 when they are added to the lexicon, which increases by 1 every time they are used for segmentation. The strongest chunks have priority during parsing and are more likely to be selected when there are multiple ways the model can legally segment the utterance. Unlike PARSER, the activation levels in PUDDLE do not decrease through decay or interference. Instead, they represent the total number of times each chunk has been used to process the input.

Like these other accounts, each chunk in CIPAL has a strength that denotes the model’s familiarity with the sequence. But unlike these previous models, chunk strength in CIPAL is represented with processing times (e.g., 200 ms), as we assume that each chunk has a processing cost that determines the time it takes to recognize the corresponding pattern in the input. By using these timings, we can directly map changes in chunk strength to the developmental patterns observed in LWL studies. Although CIPAL does not capture every aspect of children’s lexical processing, such as the effects of semantics (e.g., Borovsky et al., 2016), these chunk processing times make it possible to explore the concurrent development of word forms and speed of lexical processing within CIPAL, whereas other frameworks focus solely on word discovery without considering other processes in language development. Also, connecting the model’s chunk strength to the reaction times observed in developmental experiments introduces additional constraints on learning



and limits the researcher's degrees of freedom, compared to using arbitrary weights or activity levels with no empirical grounding.

### Assumption 3: Learning Is Constrained by Incremental Processing and a Finite STM

Languages transmit information sequentially; we do not hear all the sounds simultaneously but rather one at a time. Visual world studies have found that our language processing abilities are fast and incremental, as participants will shift their gaze in response to new information in an utterance as it unfolds in real time (e.g., Huetig et al., 2011; Tanenhaus et al., 1995). Yet, the asynchronous nature of language and comprehension is often overlooked in computational models of word discovery and vocabulary development. It is common for entire utterances, or even entire corpora, to be presented all at once and processed by the model as a single batch of data (e.g., Brent & Cartwright, 1996; Goldwater et al., 2009; Jones & Rowland, 2017).

As its name suggests, CIPAL processes and learns incrementally. The model receives language input one phoneme at a time and attempts to integrate new input with earlier material as it arrives. This incremental processing is implemented using timing parameters that determine how long critical operations within CIPAL take to complete. There are parameters that control the speech rate (160 ms), the phonological decay rate (800 ms), and the initial processing time for new phonemic chunks when they are added to LTM (1,200 ms). These specific parameter settings are explained in the Model Parameters section. The chunks in LTM also have individual processing times that get faster when they are used to recode the input.

The timings and incremental processing features in CIPAL interact to constrain the model's learning in two ways. First, since new phonemes are presented every 160 ms and decay after 800 ms, CIPAL can hold a maximum of five chunks in STM, which is consistent with the storage capacity estimates observed in memory studies (Cowan, 2001, 2010). At first, the model will be limited to storing phonemes, with STM acting as a sliding window over the input since older material is lost as new phonemes arrive. With experience, the amount of language input the model can store will increase as it builds progressively larger chunks that allow it to compress more information into STM. The input will also be active for longer since the chunks remain in STM until all their constituent phonemes decay, which means that older material is reactivated when it is integrated with more recent input. In contrast to the PUDDLE model, which builds chunks for entire utterances before discovering smaller patterns (Monaghan & Christiansen, 2010), these STM limitations force CIPAL to acquire a variety of diphones and sublexical patterns, which become the foundation for building lexical and multiword chunks. Consequently, knowledge in LTM has a hierarchical structure. CIPAL learns about the input by clustering two existing chunks together, with the phonemes of the language serving as the root nodes of the network (see the LTM section of Figure 1).

The second constraint is that CIPAL can only learn from adjacent chunks once they are both fully retrieved from LTM. In the example shown in Figure 1, the model would not learn a chunk for the word *apple* (i.e., *apəl*) until the chunks for both *ap* and *əl* have been processed. During the early stages of learning, language material will often decay from STM before it can be used to build new chunks. However, the processing times for the chunks used to recode the input will get faster even if they are not fully retrieved by

the time they decay, so repeated exposure will help the model build chunks for these patterns. As a chunk's processing time continues to get faster with experience, the model will also have more time to learn new representations before they decay. For instance, if a chunk has a processing time of 200 ms, the model has 600 ms to learn from this chunk before it decays (after 800 ms), whereas if a chunk has a 500 ms processing time, the model has only 300 ms. Collectively, this means that each chunk in CIPAL has a speed of processing that increases with experience and influences subsequent word discovery, which is consistent with results from longitudinal studies using the LWL task (e.g., Weisleder & Fernald, 2013).

Figure 1 also shows that chunks at deeper levels of the LTM hierarchy can have faster processing times than their parent nodes. This is because processing times are determined by the duration the chunks are held in STM rather than the frequency of patterns in the input. When a new chunk is created, it is initially assigned a processing time calculated from the average of its two constituent chunks (e.g., *bi* = 100 ms, *da* = 200 ms → *bida* = 150 ms). Since CIPAL recodes the input using the largest chunks in LTM, the new chunk may be used more frequently and ultimately have a shorter processing time. This is similar to the way that chunk weights are adjusted in the PARSER model (Perruchet & Vinter, 1998). However, unlike PARSER, the processing times in CIPAL do not slow down through the effects of decay and interference.

### Algorithm Description

The CIPAL architecture can be understood as a sequence of eight operations, which are repeated in cycles until the model reaches the end of the utterance (or exposure, in the case of the artificial languages used in statistical learning experiments), and the final chunk decays from STM.

1. Update the current time in the model by 160 ms (or an alternative value set by the *speech rate* parameter; see the Model Parameters section).
2. Present the next phoneme in the utterance, unless the end of the utterance (or exposure) has been reached.
3. Start retrieving the chunk for the new phoneme from LTM or create a new chunk for the phoneme if one does not exist.
4. Add the phonemic chunk to STM.
5. Attempt to learn new chunks from the units that have been fully retrieved from LTM (i.e., processing time  $\leq$  current time). This involves concatenating two adjacent and fully processed chunks stored in STM to create a new chunk in LTM (e.g., *ap*, *əl* → *apəl*). This learning process is explained in detail in Example 1.
6. Recode the contents of STM into the smallest number of units possible using the largest chunks in LTM that match the input.
7. Remove the decayed sequences from STM (i.e., decay time  $\leq$  current time). The decay times for each chunk are taken from its most recent constituent phoneme (e.g., *əl* from *apəl*). This is explained in more detail in Example 2.

8. Adjust the processing times of the chunks that remain in STM using a sigmoid function (see the Model Parameters section).

A video illustrating these processing cycles in detail is available in the additional online material (see *videos/CIPAL.mp4* at <https://osf.io/fhrxg/>).

### Example 1: Building a Chunk Hierarchy

Figure 2 illustrates how CIPAL gradually builds a hierarchy of chunks through language experience,<sup>4</sup> using a model that was repeatedly presented with the word *apple* as a sequence of phonemes (*a*, *p*, *əl*). Each panel shows the contents of CIPAL’s LTM after a set number of presentations (between 1 and 40), with the model starting with an empty LTM at the outset of each simulation. The code, results, and a video demonstration of each successive presentation is available in the additional online material (see *videos/fig-2.mp4* at <https://osf.io/fhrxg/>).

After one presentation of *apple* (Figure 2A), the model creates three new chunks for the phonemes in the sequence (e.g., *a*, *p*, *əl*). Since these phonemes are the basic perceptual elements of the language, they are added to LTM as root nodes in the chunk hierarchy with an initial processing time of 1,200 ms (see the Model Parameters section). The model then starts transferring these new chunks to STM, so they can be used to recode the input and build new representations. Since the memory traces for the phonemes decay 800 ms after their initial presentation, they become inaccessible before they are fully processed, so they are not used to construct new chunks. However, whenever a chunk is being retrieved or actively stored in STM, its processing time is adjusted using a nonlinear sigmoid function (see the Model Parameters section). Thus, the processing times for the phonemic chunks decreased from 1,200 to 1,175 ms, as they were being transferred to STM, even though their memory traces decayed before they could be used to build new chunks.

Figure 2A–2D shows the change in CIPAL’s LTM when it is presented with *apple* between 1 and 25 times. The model does not learn any new chunks during this period, but the processing times for the phonemic chunks continue to improve since they are being activated by the model each time the word *apple* is presented. After 25 repetitions, these processing times reach the threshold where they become fast enough for the model to process and store two phonemic chunks simultaneously in STM. Thus, on the 26th presentation of *apple* (Figure 2E), CIPAL creates a new chunk by combining the first two phonemes into a single unit (*a*, *p* → *ap*). The mean processing time of the two phonemes being combined (*a*, *p*) is used as the initial processing time for the new chunk. On the 27th presentation (Figure 2F), it learns the entire word by joining the chunk *ap* with the remaining phoneme *əl* (*ap*, *əl* → *apəl*). To learn these new chunks with CIPAL’s default parameter settings, the model needs to finish processing the first unit in less than 800 ms, before the memory trace decays. But it also needs to process the second unit within a shorter window of 640 ms. This is because new phonemes are presented every 160 ms and decay 800 ms later, so the model needs to retrieve the second element in less than 640 ms, otherwise the first phoneme will be lost before the second is fully processed (800 ms – 160 ms = 640 ms). This is an important constraint on learning in CIPAL: The two units being chunked need to be fully processed before the first unit decays.

In Figure 2A–2D, all three chunks in CIPAL’s LTM have identical processing times. However, these individual processing times diverge once the model starts combining phonemes into larger chunks (Figure 2E–2I), with the larger chunks reaching faster processing times than those for the primitives of the language. This is because, as new input is presented, the model searches LTM for the longest chunks matching the patterns observed in the input. The identified chunks are then used to recode and compress the input as it is transferred to STM, maximizing the amount of information the model can store in STM at any one time (see Norris & Kalm, 2021). When a chunk is used to recode the input, its individual processing time gets faster in LTM, similar to the memory reinforcement used in the PARSER model (Perruchet & Vinter, 1998). This explains why, after 40 presentations of *apple* (Figure 2I), the three fastest chunks in the model’s LTM are *a*, *ap*, and *apəl*; as the utterance unfolds one phoneme at a time, the model actively recodes the input, retrieving and reinforcing the chunks matching the patterns on the left edge of the sequence.

The reason why *apəl* is ultimately the fastest chunk in the model after 40 presentations is because it is used in the final parse of the input and remains active in STM longer than any other chunk. The shorter chunks for *a* and *ap* are only used temporarily and are displaced as new material is presented (i.e., *a* → *ap* → *apəl*). However, since *apəl* represents the entire utterance and cannot be compressed any further, this chunk is held in STM until the memory trace decays. This means that the chunks for *a* and *ap* are active for 160 ms each, whereas *apəl* is active for 800 ms and therefore receives a larger boost in processing speed. This interaction between incremental language processing and a bias for using the largest chunks matching the input also explains why the chunk for *ap* is faster than the chunk for *a*; once the model learned *ap* as a chunk (Figure 2E), it was selected over the phonemic chunks *a* and *p* and stored in STM for longer.

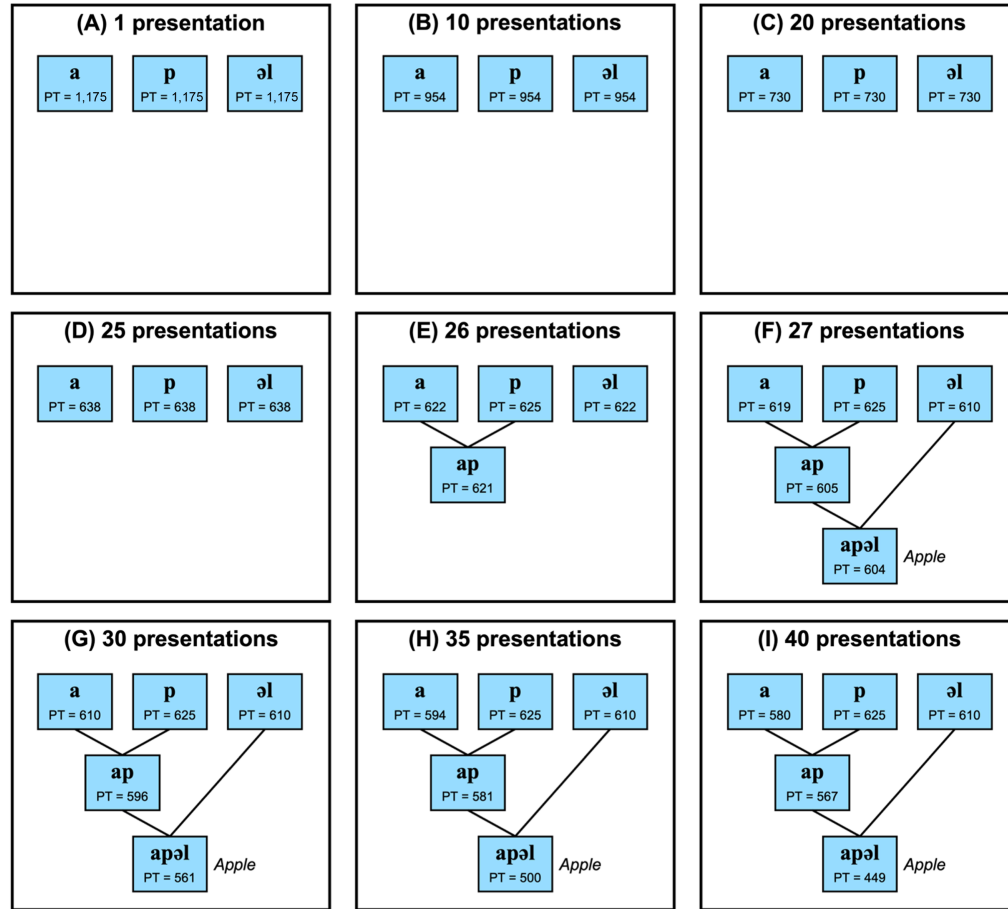
Another consequence of CIPAL’s incremental learning and processing mechanisms is that it will sometimes learn patterns in the earlier parts of the sequence first, depending on the processing times of the subchunks being combined. This is the reason why the model in Figure 2E–2I has a chunk for the first bigram in *apple* (i.e., *ap*) but not the second (i.e., *pəl*). After reaching the point where multiple phonemes could be held in STM simultaneously, the model learned the first bigram and immediately used this knowledge to recode the material into two chunks (*ap*, *əl*). The next step was to build a chunk for the entire sequence rather than learning the second bigram. Thus, the model does not learn every phonemic *n*-gram that appears in the input. Instead, it actively processes and recodes utterances as they are presented in real time and builds new chunks from the compressed representations of the input.

A video demonstrating the interaction between learning and processing in CIPAL is available in the additional online material (see *videos/CIPAL.mp4* at <https://osf.io/fhrxg/>).

<sup>4</sup> It should be noted that there were no explicit links between chunks in our implementation of CIPAL. The learning mechanism is hierarchical since larger chunks (e.g., *bida*) are formed by concatenating two smaller chunks (e.g., *bi* + *da*) already stored in LTM. However, LTM is represented as a table of chunks and processing times in the CIPAL architecture (as shown in Table 3). We visualize LTM as an explicit hierarchy in Figures 1 and 2 to help convey the hierarchical nature of CIPAL’s chunk learning mechanism.

**Figure 2**

*The Contents of Long-Term Memory for a CIPAL Model Trained With “Apple” as a Stream of Phonemes (a, p, əl) Between 1 and 40 Times*



*Note.* A video showing the changes in long-term memory after every presentation is available in the additional online material (see *videos/fig-2.mp4* at <https://osf.io/fhrxgf/>). PT = processing time; CIPAL = Chunk-Based Incremental Processing and Learning. See the online article for the color version of this figure.

### Example 2: Compressing STM With Chunks in LTM

Table 1 demonstrates how CIPAL incrementally processes language input—specifically, the child-directed utterance “It’s all gone”—by recoding the material using the chunks it has stored in LTM. The columns correspond to four separate CIPAL models, each with a different set of chunks (listed in Table 2). These four models had identical chunking mechanisms, STM capacities, and parameter settings. However, Table 1 shows that as CIPAL accumulates new chunks, it can store more input in STM and hold this information for longer before it decays. The code and results for the models shown in Tables 1 and 2 are in the additional online material (see *notebooks/cipal-examples.html* at <https://osf.io/fhrxgf/>).

In Model 1, CIPAL’s LTM only contains chunks for the phonemes, which means each element in the utterance is coded with a separate chunk in STM. However, in Model 2, CIPAL also has chunks for words and sublexical patterns. It uses this knowledge to continuously compress the contents of STM into a smaller set of chunks; at 640 ms, Model 1 uses five phonemic chunks to represent

“it’s all,” whereas Model 2 has recoded the sequence into two lexical chunks. Building on the second model, CIPAL’s LTM in Model 3 contains an additional chunk for the bigram “it’s all,” which allows the model to represent the first two words of the input with one chunk, and ultimately recode the entire utterance into two chunks (at 1,120 ms). In the final model, CIPAL has learned the entire utterance as a multiword unit and can store the full sequence as a single chunk in STM.

As well as being represented more efficiently (i.e., with fewer chunks), language input remains active in CIPAL’s STM for longer when the individual phonemes are recoded into larger units. For instance, at 1,760 ms, only the final phoneme (/n/) is stored in STM in Model 1, whereas the entire utterance is active in Model 4. This is because the decay time for each chunk is derived from the final phoneme in the unit, allowing input that would otherwise decay from STM to be reactivated when it is chunked with more recent material. In Model 4, the first phoneme in the utterance (/t/) is recoded into a new chunk five different times (/t/ → /t/ → /t/ → /t/ → /t/) and appears in STM for 1,920 ms. In

**Table 1**

*The Contents of Short-Term Memory as the Utterance “It’s All Gone” (/itsɔ lɡɒn/) Is Actively Recoded in Four CIPAL Models With Different Levels of Knowledge (Shown in Table 2)*

Time (ms)	Model 1: Phoneme	Model 2: Word	Model 3: Lexical bigram	Model 4: Full utterance
0	I	I	I	I
160	I t	(It)	(It)	(It)
320	I t s	(Its)	(Its)	(Its)
480	I t s ɔ:	(Its) ɔ:	(Its) ɔ:	(Its) ɔ:
640	I t s ɔ: l	(Its) (ɔ:l)	(Itsɔ:l)	(Itsɔ:l)
800	t s ɔ: l g	(Its) (ɔ:l) g	(Itsɔ:l) g	(Itsɔ:l) g
960	s ɔ: l g ɒ	(Its) (ɔ:l) (ɡɒ)	(Itsɔ:l) (ɡɒ)	(Itsɔ:l) (ɡɒ)
1,120	ɔ: l g ɒ n	(ɔ:l) (ɡɒn)	(Itsɔ:l) (ɡɒn)	(Itsɔ:lɡɒn)
1,280	l g ɒ n	(ɔ:l) (ɡɒn)	(Itsɔ:l) (ɡɒn)	(Itsɔ:lɡɒn)
1,440	g ɒ n	(ɡɒn)	(ɡɒn)	(Itsɔ:lɡɒn)
1,600	ɒ n	(ɡɒn)	(ɡɒn)	(Itsɔ:lɡɒn)
1,760	n	(ɡɒn)	(ɡɒn)	(Itsɔ:lɡɒn)
1,920				

*Note.* New phonemes are presented every 160 ms. These timings are internal to CIPAL and are not simulated in real time. Chunks containing more than one element are grouped in parentheses. CIPAL = Chunk-Based Incremental Processing and Learning.

comparison, the same phoneme decays after 800 ms in Model 1, since it is not chunked with any subsequent material in the input.

### Example 3: The Importance of Context Variety

Although CIPAL does not compute statistics, distributional cues still help the model discover words. In natural languages, words and morphemes are combined in different ways. This context variability provides vital information for identifying the meaningful units in the input, which has been exploited in previous chunking models (e.g., Brent & Cartwright, 1996), and it is the underlying motivation behind statistical approaches like TPs (see Saffran, Newport, & Aslin, 1996).

To show how CIPAL also uses these distributional cues, we trained the model with two different input samples. In the first sample, CIPAL received one utterance repeated 90 times: “Was it there?” The second input used the same word types as the first, but they appeared in three different utterances that were repeated

30 times each: “There it was”; “Was it there?”; and “It was there.” The only difference between the two samples was the variability in the word order, which can be detected using TPs. In the static word order condition, every within-word and between-word phoneme and syllable pair has a TP of 1.0 in both directions. In the variable word order sample, all the within-word TPs are 1.0, but the average between-word TP is approximately 0.67.

The model started with an empty LTM before it was trained with the utterances from one of the two conditions as a continuous stream of phonemes. To show how the word order variability affects the representations that CIPAL builds, the full contents of the model’s LTM after training in each condition are shown in Table 3. In the static word order condition, the model did not learn chunks for any of the words in the utterance. Instead, since there was no variety in the input, the model treated the entire utterance as a complete unit and learned only a small set of chunks, including a chunk for the entire sequence. In contrast, when trained with the variable word order sample, the model learned chunks for all three words. It also

**Table 2**

*The Contents of Long-Term Memory in the Four CIPAL Models Shown in Table 1*

Model 1: Phoneme	Model 2: Word	Model 3: Lexical bigram	Model 4: Full utterance
I	I	I	I
t	t	t	t
s	s	s	s
ɔ:	ɔ:	ɔ:	ɔ:
l	l	l	l
g	g	g	g
ɒ	ɒ	ɒ	ɒ
n	n	n	n
	It	It	It
	Its	Its	Its
	ɔ:l	ɔ:l	ɔ:l
	ɡɒ	ɡɒ	ɡɒ
	ɡɒn	ɡɒn	ɡɒn
		itsɔ:l	itsɔ:l
			itsɔ:lɡɒn

*Note.* CIPAL = Chunk-Based Incremental Processing and Learning.



**Table 3**

*The Long-Term Memory Contents of CIPAL After Training With One of Two Input Samples: (a) 90 Repetitions of the Same Utterance (“Was It There?”) With No Variability in Word Order; or (b) a Rotating Set of Three Utterances (“There It Was,” “Was It There?” and “It Was There”) Each Repeated 30 Times Using the Same Word Types as the First Sample*

Static word order		Variable word order	
Chunk	Processing time	Chunk	Processing time
w	462	ð	456
ɒ	625	eə	610
z	462	ɪ	456
ɪ	625	t	610
t	462	w	462
ð	625	ɒ	625
eə	610	z	610
wɒ	304	wɒ	453
zɪ	473	zɪ	621
tð	453	tð	621
tðeə	313	ɪt	335
wɒzɪ	192	wɒz	291
wɒzɪtðeə	120	ðeə	380
		wɒzðeə	372
		ðeəɪt	340
		wɒzɪt	381
		wɒzɪtðeə	315
		ɪtwɒz	315
		ðeəɪtwɒz	293
		ɪtwɒzðeə	320

*Note.* CIPAL = Chunk-Based Incremental Processing and Learning.

learned all three utterances as multiword chunks, as well as bigrams like “was it,” “was there,” and “it was.” This shows that CIPAL’s incremental chunking mechanism and a bias for using the largest chunks to process the input allowed it to find the sequences that tend to move around the input as complete units. Thus, the model does not need to track statistics to be sensitive to the distributional structure of the input.

## Model Parameters

### Decay Rate of 800 ms

Throughout this work, we used a fixed decay rate of 800 ms based on evidence from the mismatch negativity (MMN), a preattentive auditory event-related potential that occurs when a repetitive sound sequence is disrupted by a deviant sound that breaks the continuous pattern (Näätänen, 1992). Critically, the MMN response is only observed when the regular pattern and the deviant are presented in close temporal proximity, suggesting that they both need to be available in STM for the comparison to occur. Cheour et al. (2002) detected the MMN in newborns when the deviants (a 1,100-Hz tone) were separated from the standard stimuli (a 1,000-Hz tone) by 800 ms but not 1,500 ms, suggesting that STM could not sustain the auditory trace over the longer interval. Interestingly, the MMN can be elicited over progressively longer periods in older children and adults, which suggests that there may be maturational changes in auditory sensory memory (Bartha-Doering et al., 2015). Rather than attempting to model this developmental change, we used a fixed decay rate parameter of 800 ms, since we are simulating the earliest stages of language development.

### Speech Rate of One Phoneme Every 160 ms

Several studies have found that adult speakers adjust their articulation speed based on the age and language abilities of the listener, typically speaking at a faster pace to older children and adults compared to when they are talking to young children and preverbal infants (e.g., Ko, 2012; Narayan & McDermott, 2016; Raneri et al., 2020). To identify an appropriate speech rate for the present simulations, we calculated the average phoneme duration in the Soderstrom corpus (Soderstrom et al., 2008) from the CHILDES database (MacWhinney, 2000), a longitudinal English corpus of two American mothers talking to their preverbal infants between the ages of 6 and 13 months. We selected this corpus as it covers the developmental period where infants begin to discover the first words of their language. We accessed the transcripts for the Soderstrom corpus via the *childesr* 0.2.3 package (Braginsky et al., 2022) for the R 4.2.3 programming language (R Core Team, 2023) and phonemized all maternal input using the *eSpeak NG Text-to-Speech* (2022) speech synthesizer. Utterances with missing words or those that consisted solely of babbling sounds were excluded. In a previous analysis of articulation rate, Ko (2012) used the time stamps in CHILDES transcripts to calculate utterance durations before dividing this time by the number of words in the utterance to obtain average word durations. Following the same procedure, we used the corpus timestamps to generate average phoneme durations. Consistent with Ko’s methodology, utterances longer than 10 s were excluded to reduce the impact of positive skew from items likely to contain pauses. The resulting data set contained 21,191 maternal utterances, with median phoneme duration of 165 ms and an interquartile range of 111–276 ms. On this basis, phonemes were presented to CIPAL at

a constant 160-ms intervals. Paired with the decay rate parameter of 800 ms, this meant that five phonemes could be held in STM simultaneously before the first element was lost.

It should be noted that articulation rate in natural speech is not uniformly distributed, even within a single utterance (J. L. Miller et al., 1984). We did not attempt to implement any fine-grained timing of the phoneme presentations. We also used the same 160-ms speech rate across every simulation, including the studies with artificial languages, since our aim was to test word discovery in CIPAL across different contexts without changing the model in any way. However, we maintain that calculating a realistic speech rate from corpus data is preferable to selecting an arbitrary parameter value or performing grid search optimization, as these would increase the researcher’s degrees of freedom.

### Initial Processing Time of 1,200 ms

Across all our simulations, phonemes were used as the basic elements of language. Whenever an unfamiliar phoneme was encountered, CIPAL created a new chunk in LTM to represent it. Since phonemic chunks were not formed by combining other units and taking their average processing time, they were given an initial processing time of 1,200 ms. We selected this value based on a survey of the LWL literature (described in the Results section of Study 1), where we found that 15-month-old children needed a weighted average of 1,005 ms to shift their gaze to a named referent. We increased this to 1,200 ms, since our simulations are targeting an earlier stage of language development. This value is also 50% larger than the decay rate, which means that CIPAL needed repeated experience with the phonemes before it could use them to build new chunks.

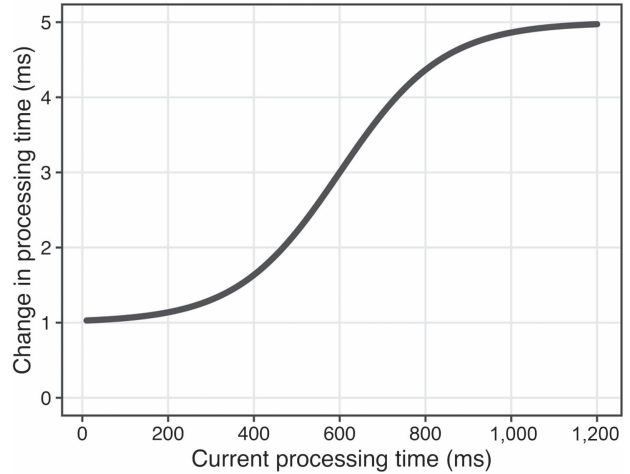
### Nonlinear Decrease in Processing Time With Experience

Whenever CIPAL uses chunks to recode the input, the processing times for these chunks get faster. Several studies have found that practice leads to nonlinear improvements in participants’ speed and accuracy across different tasks and domains (Delaney et al., 1998; Heathcote et al., 2000; Logan, 1992; Newell & Rosenbloom, 1981). Specifically, the rate of change in performance as a function of practice appears to diminish as participants reach faster speeds and higher accuracy levels. Therefore, we used a sigmoid curve to determine the magnitude of the processing time adjustments in CIPAL, which is illustrated in Figure 3. These adjustments ranged from approximately 5 ms for the slowest chunks to 1 ms for the fastest. For instance, chunks with a processing time of 1,000 ms decreased by approximately 4.86 ms, whereas those with a time of 200 ms were adjusted by around 1.13 ms. The inflection point of the curve was at 600 ms, which is half the initial processing time assigned to new phonemic chunks. At this level, the chunks were adjusted by exactly 3 ms. To prevent negative processing times, we set a floor of 10 ms, which means that all chunks in the model had a processing time between 1,200 and 10 ms.

The sigmoid curve also provides a suitable function for modeling chunk decay, where processing times become slower through lack of use. This is implemented in the CIPAL architecture using an inverted version of the curve shown in Figure 3, where the fastest chunks receive the largest penalty. However, our pilot studies found that even slow decay rates prevented the model from discovering new words and led to an overall decrease in mean chunk speed with experience. Monaghan and Christiansen (2010) observed similar

**Figure 3**

*The Magnitude of the Processing Time Adjustments Is Determined by the Chunk’s Current Processing Time Using a Sigmoid Function*



problems in their pilots with the PUDDLE architecture. Therefore, the decay parameter is disabled by default in the architecture, and we did not use it in any of the simulations presented in this work.

### Model Development and Implementation

CIPAL was implemented in the Julia 1.10 programming language (Bezanson et al., 2017). The architecture was developed according to the theory-driven testing methodology (Lane & Gobet, 2003, 2012), which emphasizes the use of reproducible tests to show that (a) the basic functionality works (unit tests); (b) the core theoretical processes have been faithfully implemented (process tests); and (c) key behavioral effects can be simulated (canonical results tests). The unit and process tests are provided with the CIPAL source code in an Open Science Framework repository (see *src/CIPAL* and *tests/CIPAL* at <https://osf.io/fhrxg/>). Since this is a new framework, the simulations presented in this article represent the canonical results tests. Coverage calculations from the *Coverage.jl* 1.6.0 package for Julia revealed that every line of the CIPAL code was accessed during testing (i.e., 100% coverage). We have also provided an example of how to use the CIPAL architecture within the R 4.3.2 programming language (R Core Team, 2023) via the *JuliaCall* 0.17.5 package (Li, 2019) in the additional online material (see *notebooks/cipal-juliacall.html* at <https://osf.io/fhrxg/>).

### Comparisons With Other Chunking Models

We developed the CIPAL architecture to determine whether incremental chunking could explain how children discover words in both natural and artificial languages using a learning strategy that is compatible with evidence from other areas of language development research. In Section 1: Word Discovery in Child-Directed Speech, we examine whether the model can discover words in continuous child-directed speech from 15 different natural languages. We then test whether CIPAL can replicate several influential findings from statistical learning experiments with artificial languages in Section 2: Simulating Word Discovery in Artificial Language Experiments. To

evaluate CIPAL’s performance in each domain, we also ran identical simulations with other chunking architectures.

For the natural languages in Section 1: Word Discovery in Child-Directed Speech, we compared CIPAL with the PUDDLE architecture (Monaghan & Christiansen, 2010) using the same corpora and target words. We selected PUDDLE as our benchmark as previous work has found that it is effective at locating words across different languages and that it has greater precision and recall than statistics-based approaches (Caines et al., 2019). For the experiments in Section 2: Simulating Word Discovery in Artificial Language Experiments, we compared CIPAL with the PARSE architecture (Perruchet & Vinter, 1998). While there are many other chunking architectures in the statistical learning literature (e.g., Retention and Recognition: Alhama & Zuidema, 2017; Truncated Recursive Autoassociative Chunk Extractor: French et al., 2011; Minimum Description Length Chunker: Robinet et al., 2011), we used PARSE, because it is a thoroughly tested framework that has simulated results from many statistical learning experiments. Previous studies have used PARSE to show that chunking can explain why participants can distinguish words from nonwords (Perruchet & Vinter, 1998), part words (Perruchet & Desauty, 2008; Perruchet & Vinter, 1998), phantom words (Perruchet & Poulin-Charronnat, 2012), and sublexical units (Giroux & Rey, 2009). Thus, since PARSE has already modeled many of our target effects, the purpose of running these new simulations was to obtain “apples-to-apples” comparisons with CIPAL to help contextualize and evaluate the results. The full details of our implementation and simulations with PUDDLE and PARSE are available in the Supplemental Materials.

## Section 1: Word Discovery in Child-Directed Speech

### Study 1: Corpora From 15 Different Languages

Our first study examined whether CIPAL can discover words in unsegmented transcripts of parental speech. The purpose of this study was to thoroughly road test the model on a diverse set of languages, as previous computational modeling studies have observed substantial cross-linguistic variation in performance (Batchelder, 2002; Caines et al., 2019; Fourtassi et al., 2013; Gervain & Guevara Erra, 2012; Jarosz & Johnson, 2013; Phillips & Pearl, 2014; Saksida et al., 2017). We trained CIPAL with 15 languages from the West Germanic (English, German, Dutch), North Germanic (Norwegian, Danish), Romance (French, Spanish, Portuguese, Italian), Slavic (Croatian, Czech, Serbian), Uralic (Estonian), Koreanic (Korean), and Celtic (Irish) families. We predicted that CIPAL would show a gradual growth in both vocabulary size and speed of processing in all languages but with cross-linguistic differences in growth rate.

### Child-Directed Speech Samples

Samples of child-directed speech were obtained from the CHILDES database (MacWhinney, 2000) using the *childesr* 0.2.3 package (Braginsky et al., 2022) in the R 4.2.3 programming language (R Core Team, 2023). We initially considered all 44 languages with monolingual samples in the database, but only those that met three criteria were ultimately included in the study. The first requirement was that at least one corpus with 10,000 utterances of child-directed speech was available for the language. This specific quantity was a compromise between including a variety of languages and having

enough data in each to evaluate CIPAL’s performance. Other modeling studies of word discovery have used similar corpus sizes (Batchelder, 2002; Blanchard et al., 2010; Brent, 1999; Brent & Cartwright, 1996; Caines et al., 2019; Christiansen et al., 1998; French et al., 2011; Goldwater et al., 2009; Monaghan & Christiansen, 2010; Venkataraman, 2001). For consistency and quality control, we only counted utterances directed at children up to 36 months that were produced by parents or grandparents and did not contain any missing words. The second criterion, for practical reasons, was that the language was supported by the *eSpeak NG Text-to-Speech* (2022) multilingual speech synthesizer, which we used to produce phonemic codes from the orthographic transcripts. The last requirement was that each language had an official adaptation of the MacArthur–Bates Communicative Development Inventory (CDI) Level 2 questionnaire (Fenson et al., 1994).

To reduce the potential impact of random variance, we ran simulations with every corpus that met our inclusion criteria rather than selecting only one sample per language or mixing samples from different children. In total, 70 corpora from 15 languages were included in the simulations (see Table 4). We extracted the first 10,000 utterances from each corpus and converted them into phonemic codes. Due to the large quantity of data available for English, we ran separate analyses for the American and British samples (identified as English U.S. and English U.K., respectively). A complete list of the corpora used in Study 1 is provided in the additional online material (see [notebooks/data-analysis.html](https://osf.io/fhrxg/) at <https://osf.io/fhrxg/>).

### Using the CDI to Evaluate Word Discovery in CIPAL

Previous modeling studies have designed algorithms to recover the boundaries between individual words or morphemes after they have been removed from the input (e.g., *doyouseethekitty* → *do you see the kitty*). The most widely used metrics for evaluating these models are *precision* and *recall* (e.g., Brent & Cartwright, 1996; Caines et al., 2019; Christiansen et al., 1998; Goldwater et al., 2009; Monaghan & Christiansen, 2010). Precision shows how many of the tokens, types, or word boundaries suggested by the model also appear in the corpus, while recall shows how many of the tokens, types, or word boundaries in the corpus the model could identify.

A problem with these metrics is that they penalize algorithms for using chunks that do not correspond to words in the corpus. Yet, there is substantial evidence that children use different types of chunks to segment and process their input, including sublexical morphemes (Mintz, 2013) and multiword phrases (Arnon & Christiansen, 2017; Christiansen & Arnon, 2017; Contreras Kallens & Christiansen, 2022; Theakston & Lieven, 2017). For this reason, CIPAL learns a variety of lexical, sublexical, and multiword chunks and does not always segment the input into words. Instead, the model uses the fewest and largest chunks available to maximize the amount of information stored in STM; it favors using one multiword chunk (e.g., *theshoe*) over separate chunks for the individual words (e.g., *the shoe*). This is sometimes described as *undersegmentation* (Blanchard et al., 2010; Gervain & Guevara Erra, 2012; Goldwater et al., 2009; Pearl et al., 2010), which implies that the model has failed to learn enough information to break down the utterances into words. Clearly, this is not the case in CIPAL since it starts with chunks for individual phonemes and builds progressively larger units. Multiword chunks are a sign of maturity in the model. They are often formed by

**Table 4***Input Characteristics of the Corpora Used in Study 1*

Language	<i>N</i>	Age	No. of items on CDI	% of CDI items in corpus	Token	Type	% One-word utterances
Croatian	1	17;3–32;8	723	53.5	37,521	3,555	25.8
Czech	1	19;5–32;0	544	59.6	36,908	4,728	22.8
Danish	2	12;26–34;21	713	56.5	37,001	1,926	31.2
Dutch	6	17;9–34;14	888	49.4	43,118	2,546	19.2
English (U.K.)	16	20;22–29;6	701	64.3	39,278	1,862	22.3
English (U.S.)	17	5;30–35;17	668	68.8	40,295	2,086	20.5
Estonian	1	19;24–24;26	657	55.7	41,190	4,564	20.5
French	8	11;17–35;21	682	58.8	43,094	2,880	22.9
German	4	10;1–34;27	580	67.5	44,942	3,035	23.5
Irish	1	17;8–21;28	911	44.6	49,224	2,862	13.0
Italian	1	17;4–25;11	693	65.1	46,346	3,562	13.9
Korean	3	15;8–31;0	638	59.2	37,910	6,915	19.3
Norwegian	1	24;2–27;22	706	62.9	53,354	3,660	16.3
Portuguese	2	17;9–32;9	657	60.3	45,527	2,148	15.3
Serbian	2	18;2–29;23	538	34.9	38,218	3,889	20.5
Spanish	4	11;1–28;16	601	63.3	37,226	2,542	22.8
Pooled data	70	5;30–35;21	686 (81)	61.6 (8.5)	40,934 (6,391)	2,650 (1,250)	21.3 (6.1)

*Note.* Each sample contained 10,000 utterances. *N* represents the number of individual corpora run for each language based on the availability of data meeting the inclusion criteria. Age shows the age range of the target children across all samples in *months; days* format. Corpus statistics are means (and standard deviations) across all samples for each language. CDI = Communicative Development Inventory.

combining two or more word-level chunks into a single unit, which allows the model to hold more of the input in STM for longer.

Instead of evaluating CIPAL using precision and recall, we used the vocabulary checklists from the CDI questionnaires (Fenson et al., 1994) and measured how many of the items were represented as single chunks in the model’s LTM. The CDI is a questionnaire given to caregivers containing a checklist of words that are likely to appear in children’s vocabulary at different ages. The caregivers are instructed to mark all the words their child can understand but not yet say and separately mark the words they can both understand and say on their own. This provides measures of receptive and expressive vocabulary, respectively. The CDI is a versatile tool that has helped researchers to estimate vocabulary norms across different languages and populations (see the *Wordbank* database; Frank et al., 2017).

There are over 100 adaptations of the CDI targeting different languages and populations, including all 15 languages meeting our inclusion criteria. These CDIs are called *adaptations* rather than *translations*, because it is not possible to use the exact same words for every language. For example, some of the function words that appear on the original American English CDI cannot be directly translated into other languages where these concepts are communicated differently. Also, many of the common nouns, such as the words for food items, are not universally relevant and will be less familiar to infants from different cultures. These words are often carefully substituted to ensure that the same information is captured. For instance, *peanut butter* is on the original American English CDI, and *aceituna* (i.e., *olive*) is on the European Spanish CDI, but not vice versa. Although the specific items may vary depending on the language and culture of the target population, the core structure of the CDI is consistent, as each adaptation includes the same semantic categories (e.g., actions, clothing, places) and has a similar balance of nouns, verbs, adjectives, and other words. Thus, the CDI provides a list of words calibrated for the specific population being studied.

In this study, we used the vocabulary checklists from the CDI-II, which is typically used to measure vocabulary size in children between

15 and 30 months. We used the CDI-II rather than the CDI-I (which is designed for children aged 8–18 months), because it has a longer checklist that includes most of the items featured on the CDI-I. We obtained the word lists for 11 languages from Wordbank: Croatian, Czech, Danish, English (U.S.), French (France), German, Italian, Korean, Norwegian, Portuguese (European), and Spanish (European). The word lists for the other four languages were taken directly from the original questionnaires. We also obtained a British English CDI checklist (Lincoln CDI), which we used for the models trained with British English corpora. All the CDI checklists used in this study are officially recognized by the CDI Advisory Board, and the adaptations always matched the corpus samples (e.g., British English CDI with British English corpora).

The word lists presented to CIPAL contained between 538 and 911 items. Nearly half of these words were nouns (47%) from various semantic categories, including animals (e.g., *dog*, *elephant*), food/drink (e.g., *apple*, *milk*), and clothing (e.g., *shoe*, *hat*). The lists also included verbs (16%; e.g., *bite*, *sleep*), adjectives (9%; e.g., *loud*, *fast*), and function words (12%; e.g., *what*, *they*). Variable items such as the child’s own name, their babysitter’s name, and their pet’s name were excluded. We converted the items into phonemic codes using *eSpeak NG Text-to-Speech* (2022), with the same language settings as the corresponding corpora from CHILDES.

We do not attempt to compare CIPAL to data collected with these measures in infants, as the model does not aim to capture the full spectrum of knowledge and skills that children acquire during vocabulary development. In particular, there is clear evidence showing that the order in which children learn their first words is guided by semantic properties such as concreteness and arousal (Braginsky et al., 2019; Tardif et al., 2008). However, CIPAL has no understanding of what the chunks it is learning mean in the language. Instead, we used the CDI as a list of target words that we expect CIPAL to discover in each language, providing test materials that are independent of the corpus samples used for training.

Since the target words are not taken from the training sample, many of the items do not appear in the model’s input (38.4% on



average; see Table 4). Rather than filtering the checklists and only testing the model on familiar words that appear in the corpora, Study 2 examines CIPAL's performance with larger training samples for a smaller number of languages where the majority of the CDI items appear the model's input.

### Measuring Speed of Processing

To determine speed of processing in CIPAL, we calculated the model's mean processing time for the CDI items that appeared in LTM. Thus, speed of processing was estimated from the words that the model has already learned as chunks. This is the same procedure used in the LWL paradigm (e.g., Fernald et al., 1998), where infants' processing speed is calculated from their reaction times to familiar words only (based on parental reports). These studies have consistently found that infants with larger vocabularies are faster at responding to speech containing words that they already know (e.g., Fernald et al., 2006; Peter et al., 2019).

### Simulation Procedure

We ran 70 separate models, one for each corpus that met the inclusion criteria. At the start of each simulation, CIPAL had an empty chunk hierarchy with no knowledge of the target language, including which phonemes were used as the basic speech elements. The timing parameters were kept at their default levels and were identical for every simulation. The models were trained with each corpus once, and the utterances were presented in developmental order. After every 50 utterances, we counted the number of words on the corresponding CDI that CIPAL had learned as one chunk by searching the model's LTM. We also computed the mean processing time for each of the CDI items that were represented with a chunk and measured the total number of chunks in LTM. In total, 200 measurements of each variable were taken per model, providing high-resolution data for each simulation. CIPAL did not learn from the

CDI lists, only from the corpus input, so repeated testing did not have any impact on the results.

### Analysis Procedure

To test our hypotheses, the simulated data were analyzed using frequentist linear mixed-effects models via the *lme4* 1.1-35 package (Baayen et al., 2008; Bates, Mächler, et al., 2015) in R 4.4.1 (R Core Team, 2024). Since Study 1 used corpora from different languages, the random-effect structure always included *language* as a random intercept. As an initial specification, the full fixed-effect structure was entered as random slopes (i.e., the maximal model; Barr et al., 2013). Before inspecting the fixed-effect estimates, we conducted a parsimonious selection process to ensure that the data supported the maximal random-effects specification (Bates, Kliegl, et al., 2015). We examined the random-effects variance-covariance matrix and conducted a principal components analysis to identify instances of singularity and overparameterization. Where necessary, slopes were removed until the random-effects specification was supported by the data, but we did not remove any fixed-effects during this selection process. Finally, *p* values were computed using Satterthwaite's method via the *lmerTest* 3.1-3 package (Kuznetsova et al., 2017), which has been shown to produce accurate Type 1 error rates (Luke, 2017). The code and output for all our analyses are available in the additional online material (see *notebooks/data-analysis.html* at <https://osf.io/fhrxg/>).

### Results: Word Discovery

Since the number of items on the CDI varied between languages (see Table 4), the dependent variable of our analyses was the proportion of the corresponding CDI checklist that CIPAL learned as a chunk. Table 5 shows the average CDI scores for each language after all 10,000 utterances were presented to the model. Across all 70 simulations, CIPAL acquired chunks for 42.4% ( $SD = 7.6\%$ ) of

**Table 5**

*The Performance of CIPAL and PUDDLE in Study 1 After All 10,000 Utterances Were Presented*

Language	CIPAL			PUDDLE		
	% of CDI chunked	Mean CDI processing time	No. of chunks learned	% of CDI chunked	Mean CDI activity level	No. of chunks learned
Croatian	42.9	138	53,298	19.8	68	4,159
Czech	36.6	118	59,401	17.6	104	4,275
Danish	31.5	138	33,306	27.3	94	3,411
Dutch	34.4	120	54,542	27.8	78	3,707
English (U.K.)	45.3	147	36,008	29.5	65	3,794
English (U.S.)	47.8	143	37,697	32.4	64	3,889
Estonian	40.8	131	72,140	23.7	53	4,019
French	38.8	142	43,372	17.8	82	3,934
German	44.3	122	63,302	35.5	86	4,105
Irish	35.3	123	53,306	18.4	81	4,708
Italian	38.7	114	73,683	24.0	78	4,872
Korean	56.8	111	82,269	25.2	89	4,350
Norwegian	43.9	122	69,449	28.8	72	5,730
Portuguese	34.3	139	50,985	28.4	55	3,453
Serbian	21.2	133	55,384	9.3	10	4,420
Spanish	38.3	123	48,116	20.7	90	2,288
Pooled data	42.4 (7.6)	136 (17)	46,347 (17,819)	27.1 (6.6)	71 (19)	3,850 (863)

*Note.* Statistics show the means (and standard deviations) across all simulations for each language. CIPAL = Chunk-Based Incremental Processing and Learning; PUDDLE = Phonotactics from Utterances Determine Distributional Lexical Elements; CDI = Communicative Development Inventory.

the CDI checklists, with substantial cross-linguistic differences in performance. The highest scores were in Korean, where over half of the CDI checklist was represented as a chunk in LTM ( $M = 56.8\%$ ). The lowest scores were in Serbian, where CIPAL learned less than a quarter of the checklist on average ( $M = 21.2\%$ ).

Figure 4 shows the proportion of the CDI that CIPAL discovered throughout the simulations. The plot suggests that the model continuously discovered new words in the input, but this growth rate slowed down as it acquired chunks for a larger percentage of the CDI. To confirm this trend, we performed a frequentist growth curve analysis (Mirman, 2014; Mirman et al., 2008). The fixed-effect structure contained orthogonalized linear and quadratic slopes for the number of utterances presented. The data supported the maximal random effects specification, with *language* as a random intercept and the linear and quadratic growth curves as correlated random slopes. The results confirmed a linear increase in the proportion of CDI items chunked by CIPAL as more utterances were presented,  $\beta = 1.31$ ,  $t(16.1) = 24.5$ ,  $p < .001$ , paired with a quadratic deceleration in this growth rate as training progressed,  $\beta = -0.36$ ,  $t(16.0) = 12.7$ ,  $p < .001$ .

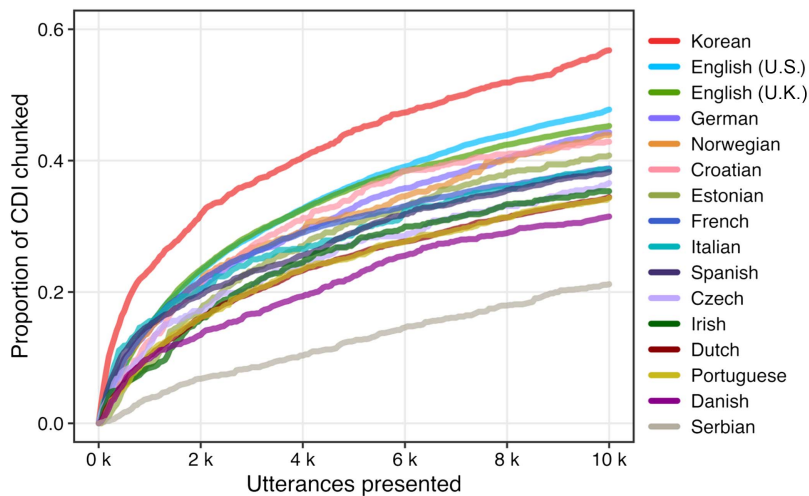
### Results: Processing Times

In our second analysis, we assessed whether CIPAL's average processing times for the chunked CDI items changed with experience. Many studies using the LWL procedure have found that children become faster at responding to familiar words as they get older (Fernald et al., 1998, 2006; Peter et al., 2019), although the shape of these growth curves varies between studies. To obtain reliable data for evaluating our simulations with CIPAL, we surveyed the speed of processing literature and extracted the mean reaction times reported in different experiments.

We conducted a systematic search using Google Scholar by screening all of the peer-reviewed journal articles that cited Fernald et al. (1998) or Fernald et al. (2006) with the words “infant” or “infancy” in the title, key words, or abstract. We chose these articles as they were the first to investigate developmental changes in children's processing speed and its relationship to vocabulary growth using the LWL task. Our two searches produced 330 and 373 citations, respectively. Each result was evaluated against several criteria. First, we only included studies that used the LWL paradigm (also called intermodal preferential looking). In these tasks, images of two familiar objects (e.g., a dog and a baby) are presented on-screen, followed by an utterance that names one of the objects with a familiar noun (e.g., “look at the dog”). The infant's looking behavior is recorded using either a video camera or an eye tracker. For this analysis, we used the average latency in the infant's first gaze shift to the named target picture (in milliseconds) as the measure of processing speed. Second, to maximize the homogeneity of the data, we only included samples of infants that were typically developing, born full term, and were learning English in a monolingual environment. Finally, we only included samples from participants aged between approximately 14 and 30 months ( $\pm 1$  month). This lower boundary was based on experimental work showing that LWL times from very young children ( $< 14$  months) are often unreliable due to their limited vocabulary (Fernald et al., 2008). We set the upper boundary because more complex stimuli are often used to study lexical processing in older infants (e.g., Peter et al., 2019).

The systematic search identified 35 samples from 16 articles, representing data from 1,706 infants between the ages of 15 and 31 months. These data are available in our additional online material (see *data/1-raw* at <https://osf.io/fhrxg/>). The sample effects are plotted in Figure 5A, which shows that children get faster at responding to their input over time. It also suggests that this relationship is nonlinear; the children appear to show a slower rate of improvement as they get

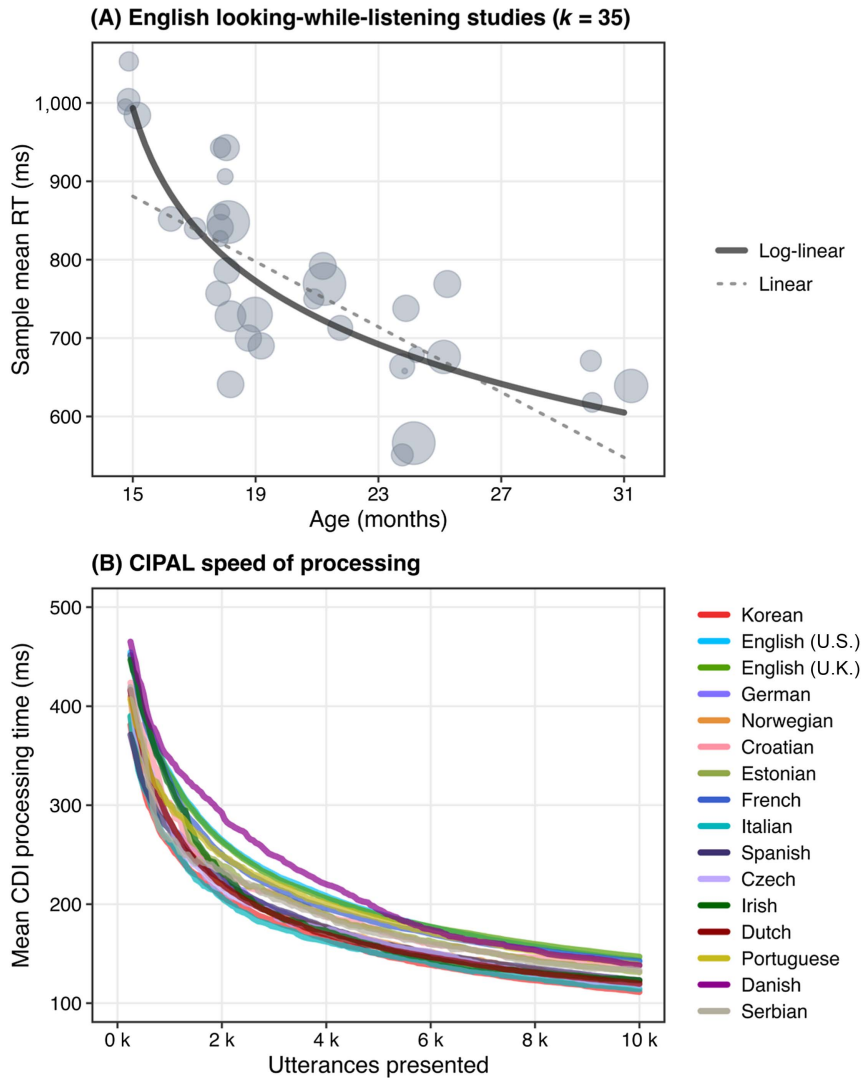
**Figure 4**  
The Proportion of CDI Items Discovered by CIPAL for Each Language in Study 1



*Note.* The legend is sorted by the proportion of CDI items discovered in each language after training with 10,000 utterances. CDI = Communicative Development Inventory; CIPAL = Chunk-Based Incremental Processing and Learning. See the online article for the color version of this figure.

**Figure 5**

*Developmental Changes in Speed of Processing in (A) English-Learning Children and (B) CIPAL*



*Note.* (A) The average reaction times observed in looking-while-listening studies with English infants between 15 and 31 months. The points are jittered along the  $x$ -axis to avoid overplotting. Their size represents the study's sample size (16–118 participants), with larger points indicating larger samples. The lines represent weighted linear and log-linear regression curves. (B) CIPAL's average processing time for known CDI times in each language. The legend is sorted by the proportion of CDI items discovered in each language after training with 10,000 utterances. CIPAL = Chunk-Based Incremental Processing and Learning; CDI = Communicative Development Inventory; RT = reaction time. See the online article for the color version of this figure.

older, with a plateau around their second birthday. This is consistent with evidence from previous work suggesting that the benefits of practice and experience diminish as participants become faster and more accurate in specific contexts (Delaney et al., 1998; Heathcote et al., 2000; Logan, 1992; Newell & Rosenbloom, 1981).

To test this nonlinear effect in the LWL studies, we compared the goodness-of-fit of two mixed-effects models. Unlike our previous analyses, the data in the model were weighted so that studies with larger sample sizes had a greater influence on the estimates. Since

some of the studies in the survey provided effect sizes at multiple age points, we included *study* as a random intercept. In the first model, the fixed-effect structure included *age* (in months) as a continuous linear variable. Before fitting the model, we subtracted 14 from the infant's true age so that the intercept represents performance at age 14 months (rather than at age 0 months), given the parameters of our systematic search. In this first model, the *age* estimate shows the change in the infant's reaction times for every 1 month increase in age. In the second model, the *age* variable was log-transformed (using the

natural logarithm) to estimate the change in the infant's processing speed with every 1% increase in age, which allows for a nonlinear rate of improvement over development. No transformations were applied to the dependent variable, which was the weighted sample mean reaction times. In both models, *age* (or *Log-age*) was retained as a random slope for *study*. Without checking the fixed-effect estimates, we compared the models on several goodness-of-fit measures. These checks found that the log-linear slope for age provided a stronger fit to the data (Akaike information criterion = 407; Bayesian information criterion = 413; root-mean-square error = 48.1) than the linear slope (Akaike information criterion = 418; Bayesian information criterion = 424; root-mean-square error = 51.9). The results of this log-linear model showed that infant's speed of processing improved by 1.32 ms with every 1% increase in age,  $\beta = -132$ ,  $t(30) = 8.47$ ,  $p < .001$ , confirming that children become faster in the LWL task as they get older, but the rate of this improvement decelerates throughout development. On this basis, we used a nonlinear improvement in processing time as the target for our simulations with CIPAL.

Figure 5B shows how CIPAL's average processing time for chunked CDI items changed with experience in the 15 languages. The patterns show a strong resemblance to the curve in the infant data. Since the processing times in CIPAL are adjusted on a sigmoid curve, the model showed larger improvements during the earliest stages of training, which gradually decreased throughout the simulations. To confirm this pattern, we fit a mixed-effects model to the data. We truncated the first 250 utterances from the data since the model's vocabulary size was too small to provide stable estimates of its average processing speed. The fixed-effects structure contained the number of utterances presented as a continuous predictor, which was log-transformed using a natural logarithm. The data supported the maximal random-effects structure, with *language* as a random intercept and *Log-utterances* as a random slope. Consistent with the growth curves observed in our survey of LWL experiments, the results showed that CIPAL's average processing time for chunked CDI items decreased by approximately 0.76 ms for every 1% increase in the number of utterances presented,  $\beta = -75.8$ ,  $t(16) = 40.0$ ,  $p < .001$ .

### Model Comparisons With PUDDLE

The statistics in Table 5 shows that CIPAL discovered a larger percentage of the CDI than PUDDLE in all 15 languages. PUDDLE acquired chunks for 27.1% ( $SD = 6.6\%$ ) of the CDI checklists on average. The highest scores were in German, where 35.5% of the CDI checklist appeared in the model's lexicon. As with CIPAL, the lowest scores were in Serbian, where PUDDLE learned 9.3% of the target words. We compared the developmental performance of both architectures using growth curves in a mixed-effects model. The fixed-effect structure contained orthogonalized linear and quadratic slopes for the number of utterances presented, crossed with *architecture* as an effect-coded factor (CIPAL = 0.5; PUDDLE = -0.5). The data supported the maximal random effects specification, with *language* as a random intercept and the full fixed-effects structure as correlated random slopes. The results confirmed that CIPAL reliably chunked more of the CDI words than PUDDLE across the simulations,  $\beta = 0.10$ ,  $t(16.0) = 8.0$ ,  $p < .001$ . The proportion of CDI items discovered by both frameworks increased linearly throughout the simulations,  $\beta = 1.04$ ,  $t(16.1) = 21.9$ ,  $p <$

.001, although the quadratic component was significant as well, showing a deceleration in growth rate over time,  $\beta = -0.29$ ,  $t(16.1) = 13.9$ ,  $p < .001$ . These growth curves also interacted with *architecture*, as CIPAL produced faster linear growth,  $\beta = 0.50$ ,  $t(15.9) = 9.1$ ,  $p < .001$  and quadratic deceleration,  $\beta = -0.11$ ,  $t(15.8) = 4.4$ ,  $p < .001$  than PUDDLE.

Despite using different chunking strategies, CIPAL and PUDDLE produced similar cross-linguistic patterns; both frameworks showed high CDI scores in English, German, and Norwegian, but the lowest scores in Serbian. Therefore, we ran a Pearson's correlation on the CDI scores generated by the two architectures, which showed a significant positive relationship,  $r = .64$ ,  $t(68) = 6.82$ ,  $p < .001$ . This suggests that some of the cross-linguistic differences observed with CIPAL were not due to the idiosyncrasies of the architecture but may reflect typological differences that affect the overall difficulty in discovering words in different languages. Similar observations have been made in previous work comparing different models of word discovery (e.g., Fourtassi et al., 2013).

### Discussion

Study 1 demonstrated that CIPAL can discover words in unsegmented corpora from multiple languages and simulate the developmental changes in children's lexical processing abilities. In particular, the growth curves for the model's processing speed showed a strong resemblance to the decrease in processing times observed in LWL studies across different age groups. The model also chunked more of the target words than PUDDLE, which suggests that incremental chunking is a more effective word discovery algorithm than a "starting big" strategy, where chunks for entire utterances are learned first and then used to find smaller lexical chunks (Arnon, 2021). Taken together, these results suggest that an incremental chunking process could explain how children discover words in their input and become more efficient at recognizing words over development.

Across all 70 simulations, CIPAL learned chunks for 42.4% of the words from the CDI checklists on average. At first glance, it might appear that the model performed poorly and failed to learn over half of the target words. However, on average, only 61.6% of the CDI items appeared in the training samples (see Table 4), which means that CIPAL did not have the opportunity to build chunks for over 38% of the words. This could explain why the simulations with Serbian showed lowest CDI scores; CIPAL built chunks for 21.2% of the checklists, but only 34.9% of the CDI items appeared at least once in the corpora. It is likely that CIPAL would continue to discover new words and acquire a larger proportion of the CDI checklists when trained with more language input. We test this hypothesis in our second study, where we used larger input samples for a smaller number of languages.

### Study 2: Large Corpora of English, German, French, and Serbian

In our first study, CIPAL learned 42.4% of the CDI on average after processing 10,000 utterances of child-directed speech. Study 2 extends these findings by testing whether CIPAL would continue to find new words and reach higher CDI scores with larger training samples. We increased the input size to 200,000 utterances for a smaller set of languages: English (U.S.), German, French, and Serbian. CIPAL discovered less than 50% of the CDI for each of



these languages in our first study (see Table 5), with Serbian showing the slowest growth and the lowest scores out of all 15 languages ( $M = 21.4\%$ ,  $SD = 5.0\%$ ).

### Input and Simulation Procedure

Study 2 followed the same simulation procedure as the previous study. The corpora were obtained from the CHILDES database (MacWhinney, 2000) using the same inclusion criteria as Study 1, except that we raised the input requirements to at least 200,000 utterances per language (not per corpus). For each language, we pooled all the data meeting the inclusion criteria into a single training set and retained the first 200,000 utterances. This meant that we ran one simulation for each language rather than separate simulations for each individual corpus. To reach the 200,000 utterances requirement for Serbian, we raised the age limit from 36 to 48 months and included input produced by the investigator. We made these exceptions since the Serbian model showed the lowest CDI scores in Study 1, making this language a strong test of whether CIPAL can discover most of the target words with more input. Table 6 provides a summary of the training data, Table 7 shows descriptive statistics for the simulation results, and Figure 6 shows the change in the model's performance over time.

### Results: Word Discovery

On average, CIPAL acquired chunks for 82% of the CDI checklists after all 200,000 utterances were presented (see Table 7). The highest scores were in German, where the model learned 92.1% of the items. Serbian still produced the lowest scores, but the model learned 73.2% of the CDI checklist, compared to 21.2% in Study 1.

The growth curves in Figure 6A suggest that CIPAL continuously discovered new words in all four languages, but the rate of improvement decelerated as the model reached higher CDI scores. To confirm these trends, we ran a growth curve analysis following the same procedure as the previous study. The fixed-effect structure contained orthogonalized linear and quadratic slopes for the number of utterances presented. The data supported the maximal random-effects structure, with *language* as a random intercept and the full fixed-effect specification as correlated random slopes. The results confirmed that the proportion of CDI items chunked by CIPAL linearly increased as the model received more input,  $\beta = 8.68$ ,  $\chi^2(1) = 5.01$ ,  $p = .025$ , but there was also a quadratic deceleration in this growth rate throughout the simulation,  $\beta = -3.42$ ,  $\chi^2(1) = 13.11$ ,  $p < .001$ .

### Results: Processing Time

Figure 6B shows the change in CIPAL's average processing time for the chunked CDI items with experience. The curves suggest a more severe deceleration and plateau in the model's processing times than in its CDI scores. Since the models were trained with large input samples, many of its chunks for the CDI items reached fast processing times or hit the floor level of 10 ms, so further experience would not produce big changes in processing speed. To confirm these trends, we fit a mixed-effects model to the data. Consistent with the analysis from Study 1, we truncated the first 250 utterances from the data since the model's vocabulary size was too small to provide stable estimates of its average processing speed. The fixed-effects structure contained the number of utterances presented as a continuous predictor, which was log-transformed using a natural logarithm. The data supported the maximal random-effects structure, with *language* as a random intercept and *Log-utterances* as a random slope. The results showed that CIPAL's average processing time for chunked CDI items decreased by approximately 0.42 ms for every 1% increase in the number of utterances presented,  $\beta = -42.5$ ,  $t(4) = 28.8$ ,  $p < .001$ .

### Model Comparisons With PUDDLE

Following the same process as Study 1, we ran identical simulations with the PUDDLE model using the same corpus samples and CDI checklists presented to CIPAL. The results of these simulations are shown in Table 7. We found that CIPAL continued to discover more words on the CDI than PUDDLE across all four languages. PUDDLE acquired chunks for 42.1% of the target words on average. The highest scores were in English, where the model learned 58.4% of the CDI. However, the model only discovered 22.5% of the items in Serbian after training with 200,000 utterances.

We compared the growth curves of the two architectures using a mixed-effects model. The fixed-effect structure contained orthogonalized linear and quadratic slopes for the number of utterances presented, crossed with *architecture* as an effect-coded factor (CIPAL = 0.5; PUDDLE = -0.5). The random-effects structure supported by the data contained *language* as a random intercept with the linear and quadratic terms for *utterances presented* as correlated random slopes. The results confirmed that CIPAL reliably learned a larger proportion of the CDI than PUDDLE across the simulations,  $\beta = 0.31$ ,  $t(32,000) = 633$ ,  $p < .001$ . The number of CDI items chunked by both frameworks increased linearly with additional input,  $\beta = 6.3$ ,  $t(4) = 22.5$ ,  $p < .001$ , with a quadratic deceleration in this growth rate over time,  $\beta = -2.6$ ,  $t(4) = 9.5$ ,  $p < .001$ . These growth curves also interacted with

**Table 6**  
*The Sample Characteristics for the Corpora Used in Study 2*

Language	Age	No. of items on CDI	% of CDI items in corpus	Token	Type	No. one-word utterances
English (U.S.)	3;0–13;13	668	92.5	813,176	9,762	22.7
French	11;17–33;15	682	82.7	881,409	15,230	20.9
German	6;13–29;0	580	97.6	896,785	19,801	22.0
Serbian	17;11–48;5	538	84.6	736,192	22,730	22.7
Pooled data	3;0–48;5	617 (69)	89.3 (7)	831,890 (73,421)	16,881 (5,661)	22.1 (0.9)

*Note.* The statistics are based on one combined input set per language. All samples contained 200,000 utterances. Age refers to the target children in the corpora and is presented in *months;days* format. CDI = Communicative Development Inventory.

**Table 7***The Performance of CIPAL and PUDDLE in Study 2 After All 200,000 Utterances Were Presented*

Language	CIPAL			PUDDLE		
	% of CDI chunked	Mean CDI processing time	No. of chunks learned	% of CDI chunked	Mean CDI activity level	No. of chunks learned
English (U.S.)	85.2	23	825,581	58.4	893	14,973
French	77.4	32	964,680	34.2	1,080	11,874
German	92.1	25	1,353,777	53.3	1,439	14,446
Serbian	73.2	35	1,049,503	22.5	42	15,420
Pooled data	82.0 (8.4)	29 (6)	1,048,385 (223,542)	42.1 (16.7)	863 (593)	14,178 (1,587)

*Note.* The statistics are based on one simulation per language. CIPAL = Chunk-Based Incremental Processing and Learning; PUDDLE = Phonotactics from Utterances Determine Distributional Lexical Elements; CDI = Communicative Development Inventory.

*architecture*, as CIPAL produced faster linear growth,  $\beta = 4.82$ ,  $t(32,000) = 157$ ,  $p < .001$  and quadratic deceleration,  $\beta = -1.60$ ,  $t(32,000) = 51.9$ ,  $p < .001$  than PUDDLE.

## Discussion

The results of Study 2 show that CIPAL can discover most of the CDI checklist in each language with sufficient training. This suggests that incremental chunking is effective at continuously discovering words in different languages. The comparatively low scores observed in Study 1 were due to a lack of opportunity rather than an inability to learn the items. With larger and more diverse input samples, there are more opportunities for the model to learn the different items as chunks. However, this does not appear to be the case for PUDDLE, which showed a much smaller improvement in performance than CIPAL and discovered less than a quarter of the target words in Serbian.

## Section 2: Simulating Word Discovery in Artificial Language Experiments

In the remaining studies, we tested CIPAL with carefully controlled artificial languages from statistical learning experiments. Experiments with artificial languages are the cornerstone of the statistical learning literature. They have provided insights into the linguistic cues and learning strategies that participants could use to identify words in continuous speech and have expanded our understanding of vocabulary development. In the canonical configuration first used by Saffran, Aslin, and Newport (1996; Saffran, Newport, & Aslin, 1996), participants listen to a continuous stream of consonant–vowel syllables (e.g., *ba, da, ku, pa, do, ti, go, la, tu, da, ro, pi*). This stream is not random but composed of a small number of three-syllable words (e.g., *badaku, padoti, golatu, daropi*) repeated throughout the exposure in a pseudorandom order. The words are concatenated into a single string that is produced in a monotonic voice at a constant tempo, leaving no pauses or other perceptual cues to identify the word boundaries. After listening to the artificial language, the participants are then tested on their ability to discriminate the words of the language (e.g., *daropi*) from either part-word foils that include a word boundary (e.g., *tudaro*) or nonword foils constructed from syllable pairs that did not appear in the exposure (e.g., *laroda*). In infants, this is typically measured using differences in their looking times in the head-turn preference procedure (Jusczyk & Aslin, 1995). The most common method in adult studies is the two-alternative forced-choice test, where a word and a foil are

presented together, and the participants select the item that sounds most like a word from the language. Studies with infants also tend to use fewer word types (e.g., four vs. six), shorter exposures (e.g., 45 vs. 300 repetitions of each type), and fewer test trials (e.g., four vs. 36 trials) than experiments with adults.

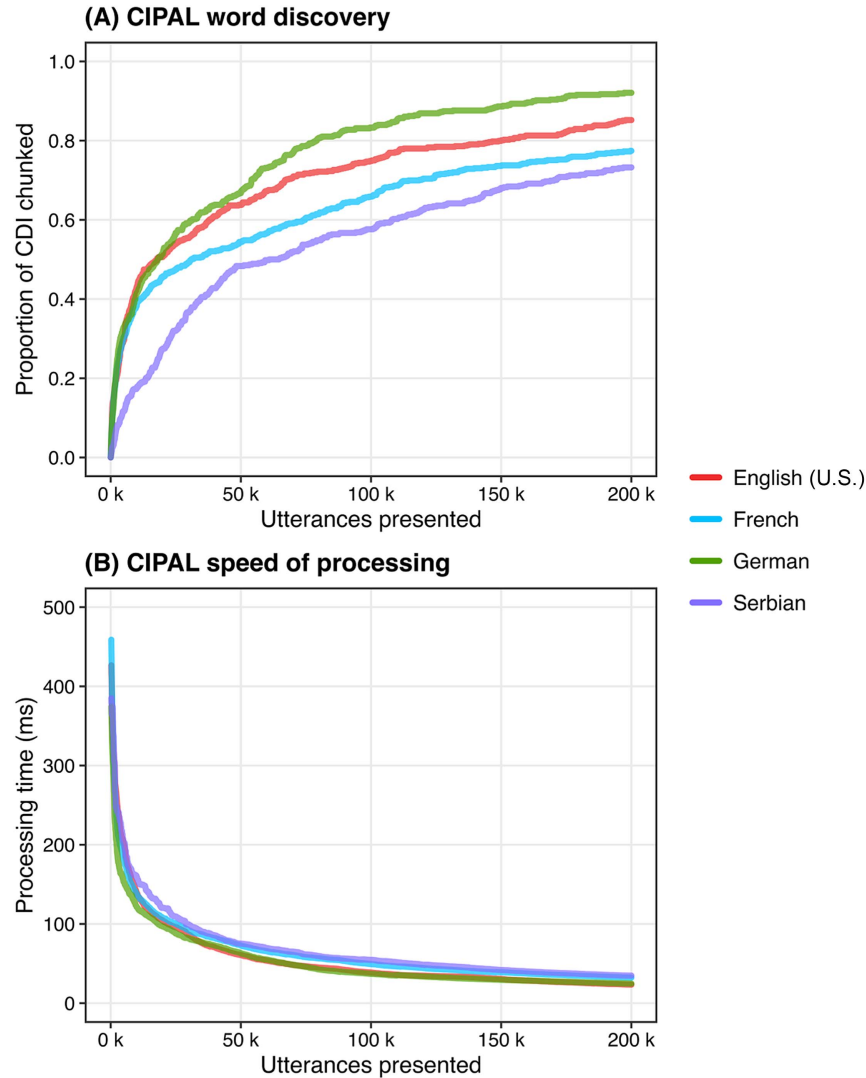
Saffran, Aslin, and Newport (1996) found that 8-month-olds can distinguish words from part-word and nonword foils. Specifically, they observed longer looking times when the infants were presented with novel foil items than familiar words that occurred 45 times in the artificial language. This appears to be a reliable effect, as a meta-analysis that aggregated 17 conceptual replications of this seminal work found a consistent novelty preference (Black & Bergmann, 2017). Similarly, in their work with adults, Saffran, Newport, and Aslin (1996) found that participants correctly identified the words with 65% accuracy for part words and 76% accuracy for nonwords. Many subsequent studies have confirmed that participants can discover words in continuous speech and have explored the boundaries of this ability by varying the characteristics of the artificial languages and test items (e.g., Endress & Mehler, 2009; Frank et al., 2010; Giroux & Rey, 2009; Isbilen & Christiansen, 2020; Kurumada et al., 2013; Perruchet & Desauty, 2008; Perruchet & Poulin-Charronnat, 2012; Perruchet & Tillmann, 2010).

In this section, we will demonstrate that several results from statistical learning studies can be simulated in CIPAL. Rather than targeting the findings of specific experiments, we aimed to simulate influential effects in the literature, some of which have been replicated multiple times with different samples. We focused on results that address the nature of the learning mechanisms that support word discovery, either by contrasting different theories or by controlling particular cues. Specifically, we test whether the model can reliably discriminate words from nonwords, part words, frequency-matched part words, phantom words, and sublexical units. Thus, while the simulations in Studies 1 and 2 demonstrate that CIPAL can extract words in samples of natural child-directed speech, Studies 3–9 provide detailed insights into how the model achieves this. Table 8 summarises the artificial languages and the model's performance across these simulations. The results for the individual test items used in each study are available in the additional online material (see [notebooks/data-analysis.html](https://osf.io/fhrxg/) at <https://osf.io/fhrxg/>).

There were two reasons why we decided to model the behavioral patterns observed in studies with small artificial languages. First, we developed the CIPAL architecture to determine whether an incremental chunking mechanism could discover words in both natural and artificial languages using learning processes that are compatible with the broader developmental literature. Most

**Figure 6**

(A) *The Proportion of CDI Items Chunked by CIPAL and (B) the Average Processing Time for the Chunked CDI Items Across the Four Languages in Study 2*



*Note.* CIPAL = Chunk-Based Incremental Processing and Learning; CDI = Communicative Development Inventory. See the online article for the color version of this figure.

computational models focus on explaining word discovery in either large naturalistic corpora (e.g., PUDDLE) or small artificial languages (e.g., PARSE), but not both contexts simultaneously. Also, these models often make assumptions that conflict with evidence from other areas of language development; for instance, many frameworks use strategies that locate words in the input at the expense of other meaningful sequences that children learn in their language (e.g., multiword units; Bannard & Matthews, 2008).

The second reason was to address a weakness in the evaluation procedure used in the previous simulations. In Studies 1 and 2, we assessed word discovery performance by searching the model's LTM for chunks matching the words on the CDI checklists. Under this testing procedure, the optimal strategy would be to exhaustively learn every phonemic  $n$ -gram in the input. This would involve

building a vast repository of chunks containing all the words in the training material, as well as chunks corresponding to every diphone, sublexical sequence, and multiword unit that appeared at least once. It is possible that CIPAL discovered more words on the CDI than PUDDLE simply because it extracted more chunks from the input (see Tables 5 and 7). However, when tested with artificial languages, this strategy would prevent the learner from discriminating words from the part-word foils that also appear in the exposure and often differ from the words by only one syllable (e.g., *daropi* vs. *tudaro*). In some studies, the frequency of the words and part-word sequences are equal, yet participants can still reliably distinguish these items (Aslin et al., 1998; Pelucchi et al., 2009b; Perruchet & Desauty, 2008). By simulating these behavioral effects in CIPAL, we show that the model does not use a naive brute-force

**Table 8**

*Means and Standard Deviations (In Parentheses) Showing the Characteristics of the Test Items and the Performance of CIPAL and PARSER in Studies 3–9*

Test item	Co-occurrence frequencies			Transitional probabilities		CIPAL		PARSER
	Syllable	Bigram	Full	FTP	BTP	No. chunks needed	Processing time	No. chunks needed
Study 3: Words versus nonwords (four trisyllabic words)								
Word	100	100	100	1.00	1.00	1.46 (0.51)	469 (205)	1.22 (0.51)
Nonword	100	0	0	0.00	0.00	3.03 (0.17)	1,136 (131)	5.80 (0.49)
Study 4: Words versus part words (four trisyllabic words)								
Word	100	100	100	1.00	1.00	1.40 (0.51)	448 (234)	1.36 (0.75)
Part word	100	67	33	0.67	0.67	1.92 (0.65)	726 (271)	5.11 (1.02)
Study 5: Words versus part words with equal co-occurrence frequencies (four trisyllabic words)								
Word	100	100	100	1.00	1.00	1.29 (0.48)	355 (179)	1.55 (1.12)
Part word	185	142	100	0.77	0.77	1.46 (0.51)	493 (183)	5.23 (1.00)
Study 6: FTP words versus part words (nine disyllabic words)								
Word	200	100	100	1.00	0.33	1.68 (0.47)	580 (194)	1.06 (0.24)
Part word	200	33	33	0.11	0.33	2.37 (0.74)	909 (393)	3.14 (0.56)
Study 7: BTP words versus part words (nine disyllabic words)								
Word	200	100	100	0.33	1.00	1.44 (0.58)	415 (248)	1.09 (0.30)
Part word	200	33	33	0.33	0.11	2.28 (0.55)	884 (267)	3.05 (0.53)
Study 8: Words versus phantom words (six trisyllabic words)								
Word	200	100	100	0.50	0.50	1.16 (0.37)	320 (116)	1.20 (0.46)
Phantom	200	100	0	0.50	0.50	2.08 (0.27)	667 (84)	3.52 (1.12)
Part word	200	86	40	0.43	0.43	1.64 (0.61)	558 (220)	3.77 (1.08)
Study 9a: Words versus sublexical units (two trisyllabic and four disyllabic words)								
Word	50	50	50	1.00	1.00	2.17 (0.53)	1,044 (288)	1.96 (0.97)
Sublexical	50	50	50	1.00	1.00	2.58 (0.86)	1,282 (501)	3.03 (1.01)
Part word	50	10	10	0.20	0.20	3.21 (0.69)	1,732 (384)	3.70 (0.50)
Study 9b: Words versus sublexical units (two trisyllabic and four disyllabic words)								
Word	300	300	300	1.00	1.00	1.68 (0.68)	456 (224)	1.00 (0.04)
Sublexical	300	300	300	1.00	1.00	2.15 (0.93)	706 (483)	3.96 (0.30)
Part word	300	60	60	0.20	0.20	2.98 (0.81)	986 (333)	3.98 (0.13)

*Note.* CIPAL = Chunk-Based Incremental Processing and Learning; FTP = forward transitional probabilities; BTP = backward transitional probabilities.

strategy of learning every subsequence in the input, and that it can provide a plausible account of the preferences observed in statistical learning experiments.

### Study 3: Words Versus Nonwords

In our third study, we trained CIPAL with the two artificial languages from Experiment 1 of Saffran, Aslin, and Newport's (1996) study with 8-month-old infants. These languages contained four three-syllable words (e.g., *tupiro*, *golabu*, *bidaku*, *padoti*) repeated in a pseudorandom order for 2 min. During the test phase, the infants' looking times were recorded as they listened to two words from the language (e.g., *tupiro*, *golabu*) and two nonwords that never appeared in the input (e.g., *dapiku*, *tilado*). As described in the previous section, infants show longer looking times for the novel foils compared to the familiar words of the artificial language, which means they can recognize reoccurring patterns in the input without ever hearing them in isolation. This has important theoretical implications, as many of the words in children's

language input never appear as single-word utterances. Thus, we tested whether CIPAL could also distinguish words from nonwords. We predicted that the model would need fewer chunks and less time to process the words of the language.

### Training and Test Stimuli

Table 8 provides a summary of the artificial languages used in Study 3. We used the same stimuli described in the appendix of Saffran, Aslin, and Newport (1996). In their original experiment, the participants were randomly allocated to one of two counterbalanced groups that used different word lists. Both groups heard the same set of items during the test phase, although the words for Group 1 were nonwords for Group 2, and vice versa. Thus, we generated 1,000 random exposures for each word list (2,000 in total). The participants in Saffran, Aslin, and Newport's study heard each word 45 times. For our simulations, we increased the exposure length to 100 repetitions. This increase is justified considering that the human participants had 8 months of experience with their native language prior to the study,



whereas CIPAL was not pretrained with child-directed speech or the syllables of the artificial language. To allow for direct comparisons between each study in Section 2: Simulating Word Discovery in Artificial Language Experiments, we used 100 repetitions for most of our artificial language simulations (see Table 8). In each simulation, the words were randomly shuffled and concatenated into a continuous text, with the restriction that the same item was never repeated twice in a row. For consistency with our previous studies, the words and foils were phonemized using *eSpeak NG Text-to-Speech* (2022) with the American English dialect. Each syllable was phonemized individually, and the codes contained no stress markers or other prosodic cues. Since the language only used consonant–vowel syllables, the words and foils always contained six phonemes. This meant that each test item could be represented by a maximum of six chunks (one for each individual phoneme) and a minimum of one chunk (the entire sequence as a single chunk). The materials and the code used to generate the exposures are available in the additional online material (see <https://osf.io/fhrxg/>).

### Simulation Procedure

Study 3 followed a similar simulation procedure to Studies 1 and 2. Separate simulations were run for each of the 2,000 random exposures generated according to Saffran, Aslin, and Newport's design. At the outset of each simulation, CIPAL had an empty LTM, with no knowledge of the syllables or phonemes used in the language. Critically, the parameters were held at their default levels and were the same for every simulation (see the Model Parameters section), which means the models used identical learning processes to the simulations from Section 1: Word Discovery in Child-Directed Speech. Each exposure language was presented to CIPAL once. Then, we tested the model with two words and two nonwords and measured two dependent variables: (a) the number of chunks needed to represent the items and (b) the total time needed to retrieve these chunks from LTM (see the Dependent Variables section). The model did not learn from the items presented at test.

### Dependent Variables

In a typical statistical learning experiment, participants are tested on their ability to distinguish the words of an artificial language from other sequences (e.g., part words, phantom units). For instance, adults are often asked to complete a two-alternative forced-choice test, where two stimuli (e.g., a word and part word) are presented together and the participants select the sequence with the greatest resemblance to the material they heard during the exposure phase (e.g., Saffran, Newport, & Aslin, 1996). Similarly, studies with infants often measure differences in the participants' looking times while they listen to the words and foil sequences (e.g., Saffran, Aslin, & Newport, 1996). The use of such paradigms means that the participants do not need to discover every unique word type in the artificial language and completely segment the input, as the knowledge threshold to recognize a difference between words and foil sequences is much lower. For this reason, we did not test whether CIPAL acquired chunks for the words while avoiding other sequences like in previous simulation studies (e.g., Perruchet & Vinter, 1998). Instead, we assessed whether the model demonstrated a representational preference for the words.

We evaluated CIPAL's preferences using two dependent variables. The first measure was the number of chunks the model needed to represent each test item. The core argument of the CIPAL theory is that children discover words, and other meaningful sequences in their language, by building progressively larger chunks. They then recode their input into the smallest number of units possible using the chunks they have stored in LTM. On this basis, we assume that participants would favor items that can be processed with fewer chunks. Our second dependent variable was the amount of time the model needed to process the test items, which was calculated as the sum of the individual processing times for the chunks used to represent each sequence. In CIPAL, chunks become faster when they are used to process the input, consistent with evidence that children's lexical processing times improve with age and their level of experience with the language (e.g., Peter et al., 2019; Weisleder & Fernald, 2013). Thus, we interpret a faster processing time as a stronger representation of the pattern.

### Analysis Procedure

The analysis consisted of two regression models, which were fit using the R 4.4.1 programming language (R Core Team, 2024). The code and results are available in the additional online material (see [notebooks/data-analysis.html](https://osf.io/fhrxg/) at <https://osf.io/fhrxg/>). The first analysis used a Poisson regression to assess the number of chunks CIPAL needed to represent each of the test items (between one and six chunks). The second used a linear regression fit to the total processing time CIPAL needed for each item. Both models included *item type* (word vs. nonword) as an effect-coded fixed factor, providing a centered estimate of the difference between words (−0.5) and nonwords (0.5). Since each simulation generated two data points for each condition, we initially fit mixed-effects models with *simulation* (1–2,000) as a random intercept. However, we removed the random effects specification entirely and used fixed-effects regressions, as the variance associated with the *simulation* grouping factor was extremely small and was not supported by the data. This suggests that differences in the pseudorandom ordering of the words had no meaningful impact on the model's learning. For the remaining studies in this section, we did not include any random effects in our analyses.

### Results and Discussion

Descriptive statistics for the stimuli and results of Study 3 are presented in Table 8. The results showed that CIPAL needed 2.1 times more chunks for nonwords than words ( $\beta = 0.73$ ,  $t = 45.9$ ,  $p < .001$ , Hedge's  $g = 4.16$ ) and retrieved the chunks for the words 667 ms faster than nonwords ( $\beta = 667$ ,  $t = 173$ ,  $p < .001$ , Hedge's  $g = 3.87$ ). Thus, CIPAL showed a preference for words over nonwords, consistent with the effects observed in experiments with adults and infants (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). The reason the model needed more chunks to process the nonwords was because there were no opportunities to learn these items as single chunks since they never occurred as complete sequences in the exposure (see *co-occurrence frequencies* in Table 8). Only the syllables appeared in the input (e.g., *da*, *pi*, *ku*), whereas the bigrams (e.g., *dapi*, *piku*) and the full nonwords (e.g., *dapiku*) did not. For this reason, CIPAL used three chunks to process the nonwords on 99.5% of the trials, as it could only build chunks for the syllables in

these items. This also contributed to the slower processing times for the nonwords, as CIPAL continuously compresses the input using the largest chunks in LTM, which become faster while they are stored in STM. After discovering the bigrams and words of the language, these larger chunks will be used to recode the input rather than the individual syllables.

#### Study 4: Words Versus Part Words

In Study 4, we examined whether CIPAL can distinguish words from part words using the two artificial languages from Experiment 2 of Saffran, Aslin, and Newport (1996). Part words are syllable sequences that span word boundaries in the artificial language. For example, Saffran, Aslin, and Newport tested 8-month-olds with part words built from the final syllable of one word and the first two syllables of another (e.g., words: *golatu*, *daropi*; part words: *tudaro*, *pigola*). These foils are a more stringent test than nonwords for two reasons. First, while nonwords are illegal sequences that never appear in the language, the participants hear the part-word sequences throughout the exposure. For example, whenever the word *golatu* is followed by *daropi*, the participants also hear the part word *tudaro*. Second, the part words have a closer resemblance to the words of the language, as they differ by only one syllable (e.g., *daropi* vs. *tudaro*). Despite this, infants and adults can reliably discriminate words from part words (e.g., Giroux & Rey, 2009; Pelucchi et al., 2009a; Perruchet & Desautly, 2008; Perruchet & Poulin-Charronnat, 2012; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996).

Since artificial languages are presented as a continuous stream without pauses or prosodic cues, the participants must exploit other features to distinguish words from part words. Typically, there are two features that differentiate the words from the foils. The first is the frequency of the test items in the exposure. In our simulations, the words of the language always occurred 100 times, but the part words appeared 33 times on average (see Table 8). The syllabic bigrams within each item (e.g., *tibu* + *budo* → *tibudo*) were also more frequent for words ( $M = 100$ ) than part words ( $M = 67$ ). It is likely that participants are sensitive to these differences, as there is extensive evidence to show that frequency distributions have a critical impact on language learning and processing (Ambridge et al., 2015; Bybee, 2006; Ellis, 2002; Lieven, 2010). The first words that infants produce are usually high-frequency content words, while low-frequency types are typically acquired later in development (Goodman et al., 2008).

The second feature is the statistical coherence of the items and how sounds are distributed in the language. In natural languages, syllables that belong to the same words have stronger statistical relationships, on average, than sequences that cross word boundaries (Harris, 1955; Saksida et al., 2017). TPs are a common measure for quantifying the co-occurrence of two elements in the statistical learning literature. They represent the diversity of syllables that appear before (BTPs) or after (FTP) a given syllable. A high TP means that two elements consistently occur together and rarely appear alone or with different syllables. Although CIPAL was trained with phonemically coded input in our simulations, we report TPs computed over syllables, as this is the most common way that artificial languages are constructed and described in the statistical learning literature. In the languages designed by Saffran, Aslin, and

Newport (1996), the four trisyllabic words of the language had higher TPs than part words in both directions (see Table 8).

#### Stimuli, Simulation, and Analysis Procedure

We followed the same simulation and analysis procedure described in Study 3. We generated 2,000 random exposures, alternating between the two counterbalanced languages described in Saffran, Aslin, and Newport's (1996) second experiment. Each exposure contained 100 repetitions of four three-syllable words, which were presented in a random order with the constraint that the same word did not occur twice in a row. The input characteristics and descriptive statistics for the languages are shown in Table 8.

#### Results and Discussion

The results showed that CIPAL needed 1.37 times more chunks to process the part words than the words ( $\beta = 0.31$ ,  $t = 17.8$ ,  $p < .001$ , Hedge's  $g = 0.88$ ) and the chunks used to represent the words were also 279 ms faster than for the part words ( $\beta = 279$ ,  $t = 49.3$ ,  $p < .001$ , Hedge's  $g = 1.10$ ). In 23.2% of the part-word trials, the model needed only one chunk to recode the sequence. In comparison, 61.5% of the words were represented with a single chunk. This was because there were more opportunities for CIPAL to learn and use these lexical chunks; as shown in Table 8, the words had both higher TPs and co-occurrence frequencies than the part-word foils. In the next study, we address this confound by balancing the frequency of the test items to explore whether CIPAL can identify the words based on distributional cues alone.

#### Study 5: Words Versus Part Words With Equal Co-Occurrence Frequencies

Artificial languages provide a way to isolate different cues and identify the sources of information that infants could use to discover words in their native language. In their original statistical learning experiments with adults (Saffran, Newport, & Aslin, 1996) and preverbal infants (Saffran, Aslin, & Newport, 1996), Saffran et al. compared high-TP and high-frequency words against low-TP and low-frequency foils. To eliminate this confound, Aslin et al. (1998) used artificial languages where the words and part words presented during the test phase had equal frequencies but different TPs. The languages used four trisyllabic words. Two of these words were presented 45 times (e.g., *pabiku*, *tibudo*), while the other two were presented 90 times (e.g., *golatu*, *daropi*). The part-word foils were built from the two high-frequency words by combining the final syllable of one word with the first two syllables of the other (e.g., *tudaro*, *pigola*). These high-frequency fragments allowed Aslin et al. to generate exposures where the part words occurred 45 times. By pairing these part words with the low-frequency words, they created a test set where the foils had lower TPs but the same co-occurrence frequency as the words (see Table 8).

Aslin et al. (1998) found that 8-month-olds showed a novelty preference, similar to their earlier experiments that did not control for differences in frequency. Subsequent studies have also found that adults and preverbal infants can discriminate words and foils when their frequencies are the same, but their forward or backward TPs are different (e.g., Pelucchi et al., 2009a; Perruchet & Desautly, 2008). Statistics-based theories have interpreted this result as

evidence that learners track the distributional properties of the input (e.g., Endress & Johnson, 2021), as it suggests that word discovery is guided by more than just the relative frequency of the sub-sequences. It also has important implications for chunking theories, which often use the repetitiveness of the words to distinguish them from other sequences (e.g., Perruchet & Vinter, 1998).

In Study 5, we tested CIPAL with the languages designed by Aslin et al. (1998). We predicted that CIPAL would use fewer chunks and have faster processing times for words than part words, despite their identical co-occurrence frequencies. CIPAL does not track statistics; it continuously recodes the input into the fewest and largest chunks possible using the knowledge it has accumulated. On this basis, there is a processing advantage for parsing the input into words. To illustrate, consider an artificial language containing the words *pabiku*, *tibudo*, *golatu*, and *daropi*. The first two words occur 45 times, and the last two words appear 90 times. This means we would test the model with the words *pabiku* and *tibudo* and the part words *tudaro* and *pigola*. Imagine that the model encounters this sequence:

*pabikugolatudaropitibudo*

If it discovers the four words of the language and uses these chunks to parse the input, it can represent this sequence as four chunks: *pabiku golatu daropi tibudo*. However, if the model uses the part word *tudaro* to parse the input, it needs at least five chunks: *pabiku gola tudaro pi tibudo*. This processing advantage for words would compound over the entire 270-token exposure. It is unlikely that the model would integrate the *gola* and *pi* fragments with their neighbors (e.g., *pabikugola*, *pitibudo*) without a longer exposure phase, as these sequences have low co-occurrence frequencies. However, it is possible that the high-frequency part words would deter the model from discovering the words of the language if these sequences were learned early. Experiments have found that familiarizing participants with part words before they hear the language has this effect (Poulin-Charronnat et al., 2017). Thus, in Study 5, we tested whether CIPAL would show a preference for words over frequency-matched part words.

### Stimuli, Simulation, and Analysis Procedure

We followed the same simulation and analysis process as the previous studies in this section. Our simulations alternated between the two artificial languages described in Aslin et al. (1998), which contained four three-syllable words. We generated 2,000 random exposures where the words and part words both occurred 100 times. To do this, we repeated the low-frequency words 100 times and the high-frequency words 185 times. The words appeared in a pseudorandom order with the restriction that the same words did not occur twice in a row. It was necessary to reduce the frequency difference between the low- and high-frequency words to 85%, as our search algorithm was unable to find a variety of different exposures when the high-frequency words occurred twice as often. Table 8 provides a summary of these languages and descriptive statistics for the results of the simulations.

### Results and Discussion

The results showed that CIPAL used 14% more chunks ( $\beta = 0.13$ ,  $t = 6.6$ ,  $p < .001$ , Hedge's  $g = 0.35$ ) and was 155 ms slower for part

words than words ( $\beta = 138$ ,  $t = 34.1$ ,  $p < .001$ , Hedge's  $g = 0.76$ ). Thus, despite having the same co-occurrence frequencies, CIPAL demonstrated a processing advantage for the words of the language. This is an important result as it shows that the discrimination effects observed by Aslin et al. (1998), and other studies that balance the frequency of the test items (Pelucchi et al., 2009b; Perruchet & Desauty, 2008), can be explained with a chunking algorithm that does not compute statistics. It also shows that such models do not need to depend exclusively on differences in relative frequency to discover words. Instead, an incremental process of learning and using progressively larger chunks is enough to find meaningful sequences in the input.

### Study 6: FTP Words Versus Part Words

Building on the results of the previous simulations, Studies 6 and 7 examine whether CIPAL can distinguish words from part-word foils in an artificial language where the words are constructed based on either forward or backward TPs. TPs are a direction-specific measure of the relationship between two elements (Perruchet & Desauty, 2008). The FTP represents the probability that  $X$  is followed by  $Y$ :

$$\text{FTP} = P(Y|X) = \frac{\text{frequency of } XY}{\text{frequency of } X}. \quad (1)$$

Likewise, BTPs represent the probability that  $Y$  is preceded by  $X$ :

$$\text{BTP} = P(X|Y) = \frac{\text{frequency of } XY}{\text{frequency of } Y}. \quad (2)$$

While both of these measures quantify the strength of the relationship between two elements in a sequence, their effectiveness at locating word candidates varies across languages. For instance, Saksida et al. (2017) found that FTP algorithms were between 11.8% and 12.1% more accurate than BTPs at predicting word boundaries in Italian, whereas BTPs consistently outperformed their FTP equivalents in Polish by 3.6%–10.2% (based on absolute and relative thresholds, respectively).

Although the words in statistical learning experiments typically have higher TPs in both directions, several studies have found that participants can still distinguish words from part words when only one of these measures is informative. For example, Perruchet and Desauty (2008) recruited two samples of adults and trained each group with a different artificial language. The first group listened to a language with nine disyllabic words that were constructed to have perfect FTPs (1.0) but low BTPs (0.33). The second group was presented with a mirrored version of the first language, where the order of the syllables was reversed to create strong BTPs (1.0) but low FTPs (0.33). The participants were numerically above chance at recognizing words over part words but only significantly above chance in the BTP group (BTP:  $M = 67.20\%$ ,  $SE = 5.59$ ; FTP:  $M = 60.32\%$ ,  $SE = 5.15$ ). In a second experiment that controlled for frequency differences between the words and part-word foils, the participants were significantly above chance at identifying words in both languages (BTP:  $M = 61.1\%$ ,  $SE = 5.11$ ; FTP:  $M = 66.7\%$ ,  $SE = 4.32$ ).

Developmental studies have similarly found that preverbal infants can distinguish words from part words, when their TPs are different in only one direction. In two studies, Pelucchi et al. (2009a, 2009b)

trained English-learning 8-month-old infants with controlled samples of Italian child-directed speech. The infants then heard sequences with either high or low TPs in a head-turn preference task. Regardless of whether the words were constructed from FTPs (Pelucchi et al., 2009b) or BTPs (Pelucchi et al., 2009a), the infants looked for longer at the high-TP than the low-TP items.

Collectively, this work suggests that participants can discover words with different statistical properties. In Study 6, we assessed whether CIPAL could distinguish words from part words with different FTPs but identical BTPs (see Table 8) using the artificial languages designed by Perruchet and Desautly (2008). We then reversed the TP structure of the language in Study 7 to test whether the model shows a preference for words over foils that have different BTPs but matched FTPs. Since CIPAL does not track distributional statistics, we predicted that it would use fewer chunks and have faster processing times for the words.

### ***Stimuli, Simulation, and Analysis Procedure***

We used the same simulation and analysis procedure as the previous studies. We generated 2,000 random exposures based on the materials described by Perruchet and Desautly (2008). Each language contained nine two-syllable words that were repeated 100 times in a random order. Unlike our previous studies, there was no restriction on whether a word could occur twice in a row since the original study did not enforce this constraint. Each word contained one unique syllable that did not appear in any other type (A, B, C, D, E, F, G, H, I) and a syllable that appeared in two other words (X, Y, Z). The shared syllable appeared in the word-final position (AX, BX, CX, DY, EY, FY, GZ, HZ, IZ). This meant that the first syllable of each word perfectly predicted the second syllable (FTP = 1). However, the BTP was 0.33, as the second syllable was preceded by three different elements. Since Perruchet and Desautly described their stimuli using letter codes, we built the language from a list of unique consonant–vowel syllables that were randomly mapped to one of the letters. Although we constructed the language using syllables, the exposures were presented to CIPAL as a stream of phonemes.

### ***Results and Discussion***

The results showed that CIPAL used 42% more chunks ( $\beta = 0.35$ ,  $t = 46.2$ ,  $p < .001$ , Hedge's  $g = 1.13$ ) and was 328 ms slower for part words than words ( $\beta = 328$ ,  $t = 101$ ,  $p < .001$ , Hedge's  $g = 1.06$ ). Since CIPAL discovers words through an incremental chunking process, it is not dependent on a specific distributional cue. This means that the model can separate words from part words in languages where BTPs are not an informative feature, extending the results of the previous simulations where words have stronger TPs in both directions.

### **Study 7: BTP Words Versus Part Words**

Building on the results of Study 6, we tested whether CIPAL would show a preference for words that had stronger BTPs than the part words, but identical FTPs. Like the previous study, we ran 2,000 simulations using the materials designed by Perruchet and Desautly (2008). The only difference is that the TP structure of the words is reversed, which means that the words had BTPs of

1.0 and FTPs of 0.33. To build the language, we generated a list of consonant–vowel syllables that were randomly and uniquely assigned to the following letter codes to create nine disyllabic words: XA, XB, XC, YD, YE, YF, ZG, ZH, ZI. The syllables for X, Y, and Z appeared in three words, whereas all other syllables only appeared in one word.

Compared to the words of the language, CIPAL used 57% more chunks ( $\beta = 0.46$ ,  $t = 57.9$ ,  $p < .001$ , Hedge's  $g = 1.49$ ) and was 469 ms slower to process the part words ( $\beta = 469$ ,  $t = 172.7$ ,  $p < .001$ , Hedge's  $g = 1.82$ ). Taken together, Studies 6 and 7 demonstrate that CIPAL is not dependent on the words having stronger TPs in either the forward or backward direction. This is consistent with the results of adult and infant studies, which have demonstrated that humans can learn words with a variety of different statistical compositions (e.g., Giroux & Rey, 2009; Pelucchi et al., 2009a, 2009b; Perruchet & Desautly, 2008; Perruchet & Poulin-Charronnat, 2012).

### **Study 8: Words Versus Phantom Words**

In Study 8, we tested whether CIPAL could distinguish words from phantom words, which are foils that never appear in the exposure but have identical TPs to the words of the language. Phantom units (also called *prototypes* or *illusory* units) were first introduced by Endress and Mehler (2009) to examine whether participants learn statistics or word-like units from their language input. In designing their artificial language, the authors created two phantom units to serve as prototypes for building the words. For example, the participants heard the words *tazepi*, *mizeRu*, and *tanoRu* repeated between 75 (Experiment 1a) and 600 times (Experiment 1d) during the training phase. These were based on the phantom unit *ta-ze-Ru*, with different syllable pairs from this prototype appearing in each word (*ta-ze-X*, *X-ze-Ru*, *ta-X-Ru*). By building the language in this way, with each syllable occurring in two different words, the TPs (in both directions) were consistently 0.5 for every pairwise relationship in both the words and the phantom units.

Across several experiments with a total of 161 adults (Experiments 1a–1d), Endress and Mehler (2009) did not find any preference for words over phantom words. When they introduced additional word–boundary cues, such as pauses (Experiment 3) or final syllable lengthening (Experiment 4), then a preference for the words was observed. These results are consistent with the predictions of statistics-based theories of word discovery (e.g., Endress & Johnson, 2021), which argue that children discover words using a mechanism that tracks the statistical relationships between successive syllables. Since the words and phantoms had identical TPs, the participants needed additional cues to identify the words in the input. The findings also conflict with the predictions of chunking models. For instance, PARSER and Minimum Description Length Chunker can distinguish words from phantoms in Endress and Mehler's language after a brief exposure (Perruchet & Poulin-Charronnat, 2012).

However, subsequent studies using phantom units tell a different story. Perruchet and Poulin-Charronnat (2012) used the same artificial language and test stimuli as Endress and Mehler (2009). Across three separate studies with a total of 108 participants, they found a consistent preference for words over phantom words. Visual studies using sequences of colored shapes have also observed reliable discrimination effects. For instance, Slone and Johnson (2018) found that 8-month-old infants looked



for longer when tested with familiar triplets that appeared in the visual language 50 times compared to unobserved phantom sequences with identical TPs. The same authors have similarly found that adults can accurately identify familiar triplets over novel phantoms (Slone & Johnson, 2015). On balance, most experiments have found that participants can discriminate words from phantom sequences (Ordin, Polyanskaya, & Soto, 2020; Ordin, Polyanskaya, Soto, & Molinaro, 2020; Perruchet & Poulin-Charronnat, 2012; Polyanskaya, 2022; Slone & Johnson, 2015, 2018), with a small number of studies reporting null results (Endress & Langus, 2017; Endress & Mehler, 2009). We, therefore, use the finding that participants show a preference for words over phantom words as our target result.

### ***Stimuli, Simulation, and Analysis Procedure***

We assessed CIPAL with the artificial language and test stimuli used by Endress and Mehler (2009) and Perruchet and Poulin-Charronnat (2012), following the same procedure as the previous studies in this section. The language contained six trisyllabic words built from two phantom words (*ta-ze-Ru* → *tazepi*, *mizeRu*, *tanoRu*; *fe-ku-la* → *fekupi*, *mikula*, *fenola*). We generated 2,000 random exposures where each word was repeated 100 times. The words appeared in a random order, but the same word did not occur twice in a row. We tested the model with all six words and the two phantom words. For comparison, we also tested the model with the 12 part-word foils used in the original experiments (see Appendix A of Endress & Mehler, 2009). Like the previous studies, we estimated the effect of *item type* on the model's representational preferences. In all our analyses, *item type* was coded with two centered contrasts, which compared (a) words with phantom units and (b) words with part words. The input characteristics and descriptive statistics for the simulations are shown in Table 8.

### ***Results and Discussion***

Compared to the words of the language, CIPAL needed 79% more chunks ( $\beta = 0.58$ ,  $t = 42.1$ ,  $p < .001$ , Hedge's  $g = 1.06$ ) and was 346 ms slower to process the phantom units ( $\beta = 346$ ,  $t = 103$ ,  $p < .001$ , Hedge's  $g = 0.88$ ). It also used 41% more chunks ( $\beta = 0.35$ ,  $t = 35.1$ ,  $p < .001$ , Hedge's  $g = 0.43$ ) and was 238 ms slower to process the part words than the words ( $\beta = 238$ ,  $t = 116$ ,  $p < .001$ , Hedge's  $g = 0.73$ ). Similar to the nonword foils we used in Study 3, the phantom units do not appear in the exposure. This means there are no opportunities for CIPAL (or other chunking models) to learn these sequences as single chunks. Instead, the model recoded the phantom units into two chunks on 94% of the trials (e.g., *taze Ru*), often by pairing a bigram chunk that appeared in one of the words (e.g., *taze* from *tazepi*) with a syllabic chunk from another word (e.g., *Ru* from *tanoRu*). CIPAL is also slower at retrieving these chunks, as it is more likely to recode the input using words after it has discovered these units.

### ***Study 9a and 9b: Words Versus Sublexical Units***

Statistics-based and chunking theories of word discovery often make similar predictions. For this reason, both frameworks have modeled many of the same findings in the statistical learning literature (e.g., Endress & Johnson, 2021; French et al., 2011; Perruchet,

2019). However, these theories make different predictions for whether participants can distinguish words from phantom units (see Study 8) and sublexical units (e.g., *bida* from the word *bidaku*). In both cases, statistics-based theories argue that participants will not discriminate these items since their statistical properties are identical. However, chunking theories predict that participants will distinguish the words from the foils since they extract complete units from the input rather than tracking the probabilistic links between elements. Most experiments have found evidence consistent with the predictions of the chunking accounts (Giroux & Rey, 2009; Perruchet & Poulin-Charronnat, 2012; Slone & Johnson, 2015, 2018). In the previous study, we found that CIPAL needs fewer chunks and less time to process words than phantom units. In our final study, we examined whether CIPAL shows a similar preference for words over sublexical patterns.

Several experiments have found that infants and adults can discriminate words from sublexical (or embedded) units. In their seminal study of adults, Giroux and Rey (2009) designed an artificial language with two trisyllabic words (e.g., *bidaku*) and four disyllabic words (e.g., *gola*). The participants listened to the language for either 2 or 10 min before completing a two-alternative forced-choice test with two conditions. First, they heard the disyllabic words paired with a disyllabic part word (e.g., *gola* vs. *labi*). Then, they heard two-syllable fragments from the trisyllabic words alongside a disyllabic part word (e.g., *bida* vs. *kugo*). In both cases, the participants were instructed to identify the items with the greatest resemblance to the input language. Giroux and Rey found that the participants accurately selected the words and sublexical units over the part words in both the 2- and 10-min exposure conditions. Although the effect sizes were similar in the 2-min condition, the authors observed larger discrimination effects for the words ( $M = 75\%$ ,  $SE = 3.9\%$ ) than the sublexical units ( $M = 65.2\%$ ,  $SE = 4.2\%$ ) after a 10-min exposure. This suggests that participants can recognize patterns of different lengths, but the sequences that function as complete units in the language become stronger than their subcomponents with experience. Critically, words and sublexical units had identical co-occurrence frequencies and TPs (see Table 8). The only difference between the items was whether they represented a complete word or a fragment of a word. Giroux and Rey also showed that these results can be simulated by the chunking mechanisms of PARSE, but not by the statistical computations of a simple recurrent network. Other studies have also observed a preference for words over embedded units in both adults and infants using visual stimuli (Slone & Johnson, 2015, 2018).

In our final study, we tested whether CIPAL shows a preference for complete words over sublexical units, as well as a preference for sublexical units over part words, in the artificial languages designed by Giroux and Rey (2009). Unlike the previous studies in Section 2: Simulating Word Discovery in Artificial Language Experiments, we manipulated the length of exposure phase to examine how the model's representation preferences changed with experience. Specifically, each word type was repeated 50 times in Study 9a, and 300 times in Study 9b.

### ***Stimuli, Simulation, and Analysis Procedure***

We used the same simulation and analysis procedure as the previous studies in Section 2: Simulating Word Discovery in Artificial Language Experiments. The language and test items were

based on the materials described by Giroux and Rey (2009). Each language contained two three-syllable words (e.g., *ABC, DEF*) and four two-syllable words (e.g., *GH, IJ, KL, MN*). Giroux and Rey described their stimuli using alphabetic symbols (e.g., *ABC*). To build the language from these blueprints, we created a list of consonant–vowel syllables (as phonemic codes) that were randomly mapped to different letters, with each syllable appearing in only one word. In Study 9a, we generated 2,000 random exposures where each word was repeated 50 times in a random order, with the constraint that the same word did not occur twice in a row. For Study 9b, we created another 2,000 exposures where each word type was repeated 300 times. For our analyses, we used Poisson and linear regression models to estimate the effect of *item type* on the model’s performance. Item type was coded with two centered contrasts, which compared (a) words with sublexical units and (b) sublexical units with part words. The input characteristics and descriptive statistics for the simulations are shown in Table 8.

## Results and Discussion

When trained with the short exposures in Study 9a, CIPAL needed 19% fewer chunks for words than sublexical units ( $\beta = 0.17$ ,  $t = 16.8$ ,  $p < .001$ , Hedge’s  $g = 0.96$ ) and 25% fewer chunks for sublexical units than part words ( $\beta = 0.22$ ,  $t = 23.6$ ,  $p < .001$ , Hedge’s  $g = 1.15$ ). The model was also 237 ms faster at processing words than sublexical units ( $\beta = 237$ ,  $t = 37.5$ ,  $p < .001$ , Hedge’s  $g = 1.05$ ) and 450 ms faster with sublexical units than part words ( $\beta = 450$ ,  $t = 71.1$ ,  $p < .001$ , Hedge’s  $g = 1.38$ ). For the longer exposures used in Study 9b, CIPAL needed 28% fewer chunks for words than sublexical units ( $\beta = 0.25$ ,  $t = 21.7$ ,  $p < .001$ , Hedge’s  $g = 1.01$ ) and 38% fewer chunks for sublexical units than part words ( $\beta = 0.32$ ,  $t = 32.5$ ,  $p < .001$ , Hedge’s  $g = 1.27$ ). It was also 250 ms faster at processing words than sublexical units ( $\beta = 250$ ,  $t = 43.6$ ,  $p < .001$ , Hedge’s  $g = 1.02$ ) and 280 ms faster with sublexical units than part words ( $\beta = 280$ ,  $t = 48.7$ ,  $p < .001$ , Hedge’s  $g = 1.07$ ).

CIPAL showed a processing advantage for words over sublexical units, even though these items had the same co-occurrence frequencies and TPs (see Table 8). This is because the model continuously recodes the input using the largest chunks in its LTM (e.g., *A, B, C* → *AB, C*) before attempting to learn new chunks from the recoded material (e.g., *AB, C* → *ABC*). The model does not build chunks for every phonemic  $n$ -gram in the input. Instead, it may sometimes need two chunks to represent a sublexical unit (e.g., *B, C*) but only one chunk for a word containing the same sequence (e.g., *ABC*). For example, if the model learned the word *bidaku* by chunking together *bida* and *ku*, it would not automatically have a chunk for the sublexical unit *daku*. Lexical chunks were also faster than sublexical units for similar reasons. The model recodes the input using the largest chunks in LTM, and these chunks become faster while they are stored in STM. Since the chunks for the words allow the model to compress more information, the model is more likely to use these chunks to recode the input. These processing advantages were observed even after a short exposure to the language and increased in magnitude with additional experience.

These findings are partially consistent with the results of Giroux and Rey’s (2009) original experiment with adults, which observed a preference for words over sublexical units when the participants listened to the language for 10 min, but not when they listened for 2 min. The increase in the magnitude of CIPAL’s preferences suggests

that the model gradually learns to distinguish the complete words of the language from sublexical sequences within these words, but this could be detected earlier in the model than in the participants of Giroux and Rey’s experiment. However, in experiments with visual stimuli, Slone and Johnson (2018) found that 8-month-old infants could discriminate words from sublexical units after an exposure phase containing 80 repetitions of each word. This suggests that a preference for words over sublexical units may emerge earlier than originally observed in Giroux and Rey’s experiment with adults, consistent with the predictions of CIPAL in Study 9a. Thus, additional experiments are needed to clarify when participants begin to discriminate words from sublexical units.

## Model Comparisons With PARSER

The implementation, procedure, and results for the simulations with PARSER are described in more detail in the Supplemental Materials. Table 8 shows the average number of chunks that PARSER needed to process the different test sequences from each study. The model produced similar preferences to CIPAL in Studies 3–8, showing a consistent processing advantage for words over nonwords, part words, frequency-matched part words, and phantom units. In Studies 9a and 9b, PARSER responded differently to CIPAL when the length of the exposure increased. When each word type was repeated 50 times, PARSER showed a reliable preference for words over sublexical units, and sublexical units over part words, consistent with empirical data from experiments with adults (e.g., Giroux & Rey, 2009) and infants (e.g., Slone & Johnson, 2018). However, when the exposure was extended to 300 repetitions of each type, the model showed no meaningful distinction between the sublexical sequences and part words; specifically, PARSER showed less than 1% difference in the number of chunks used in each condition. This was because sublexical chunks were pruned from memory after the model discovered the words of the language, since they were no longer being used to parse the input. However, evidence suggests that children retain sublexical chunks that help them to process and learn from unfamiliar sequences (Gathercole, 1995; Jones et al., 2007; Mintz, 2013). Thus, while Section 1: Word Discovery in Child-Directed Speech found that CIPAL was more effective than PUDDLE at finding words in natural languages, Section 2: Simulating Word Discovery in Artificial Language Experiments showed CIPAL was also more effective than PARSER at reproducing the results of different statistical learning experiments.<sup>5</sup>

In Study 8, the two frameworks also made different predictions on whether participants would find it harder to distinguish words from phantom units or part words. Although both models showed a clear preference for words over both foils, CIPAL used fewer and faster chunks to represent part words than phantom units, whereas

<sup>5</sup> It should be noted that, in their original work, Giroux and Rey (2009) found that PARSER showed a close resemblance to the human data. The authors trained PARSER with 400 syllables (short exposure) or 2,000 syllables (long exposure) using the same materials presented to the participants in their experiment. They assessed the model’s preferences by presenting a word or sublexical unit paired with a part word. PARSER selected one item from each pair, preferring items that appeared in the perception shaper with strong weights (or selecting one at random if neither was represented). Although it is possible that the use of a different procedure influenced PARSER’s effectiveness in Study 9, we maintain that CIPAL is a more robust chunking algorithm that can account for a variety of statistical learning effects across different contexts.

PARSER used fewer chunks for the phantom units than part words. It is unclear from the experimental literature which of these foil sequences participants are more likely to confuse with the words of the language. In Perruchet and Poulin-Charronnat's (2012) third study, participants were more accurate in trials comparing words with part words (Cohen's  $d = 1.04$ ) than words with phantom units (Cohen's  $d = 0.554$ ). However, in Slone and Johnson's (2018) experiments with visual stimuli, 8-month-olds showed a larger difference in their looking times when comparing words with phantom units than part-word sequences, although this difference was not statistically significant. Thus, further research is needed to test these conflicting predictions.

### General Discussion

Evidence from multiple experiments (Giroux & Rey, 2009; Perruchet & Poulin-Charronnat, 2012) and computational models (Monaghan & Christiansen, 2010; Perruchet & Vinter, 1998; Robinet et al., 2011) suggests that children discover words by extracting chunks embedded in their input (see Perruchet, 2019). To connect these accounts with the broader language development literature, we propose an integrated theory of word discovery, implicit statistical learning, and speed of lexical processing. The theory is implemented as a computational modeling architecture called CIPAL. The model receives language input as a continuous stream of phonemes and attempts to recode the material into larger units using chunks stored in its LTM. It acquires new chunks by combining adjacent elements that appear in the input to incrementally build a chunk hierarchy that represents knowledge at different levels. Each chunk also has a dynamic processing time that becomes faster with experience, consistent with evidence from speed of processing experiments (e.g., Weisleder & Fernald, 2013).

Across nine studies, we found that the automatic chunking mechanisms in CIPAL could discover words in natural languages, replicate the results of statistical learning experiments with artificial languages, and model the developmental change in children's lexical processing speed. In Section 1: Word Discovery in Child-Directed Speech, CIPAL was trained with child-directed speech from 15 different languages and tested with word lists from official adaptations of the CDI questionnaire. The model gradually discovered words in all the languages, showing cross-linguistic variation in performance. In a follow-up study where the model was trained with a larger sample of 200,000 utterances from a smaller set of languages (English, German, French, and Serbian), CIPAL extracted chunks for 82% of the target words on average. This was nearly twice the number of words that the PUDDLE model discovered when tested with the same materials. The processing times for these lexical chunks in CIPAL also produced growth curves matching the developmental trends observed in children with the LWL task. In Section 2, we found that CIPAL can model a series of influential findings from statistical learning experiments, showing greater versatility than the PARSER model. Specifically, CIPAL could reliably discriminate words from nonwords (Saffran, Aslin, & Newport, 1996), part words (Perruchet & Desauty, 2008; Saffran, Aslin, & Newport, 1996), frequency-matched part words (Aslin et al., 1998), phantom words (Perruchet & Poulin-Charronnat, 2012), and sublexical units (Giroux & Rey, 2009).

Previous chunking models of word discovery have attempted to isolate the words of the language by preventing the model from

learning or retaining other meaningful sequences (e.g., Brent & Cartwright, 1996; Goldwater et al., 2009; Perruchet & Vinter, 1998). However, CIPAL builds an extensive chunk hierarchy that includes lexical, sublexical, and multiword representations. This is consistent with studies showing that children learn chunks for different aspects of their native language, including sublexical morphemes (e.g., Mintz, 2013) and multiword phrases (e.g., Bannard & Matthews, 2008). Several theories have argued that these representations play a critical role in language development (Christiansen & Arnon, 2017; Jones & Rowland, 2017; Theakston & Lieven, 2017). For instance, simulations with CLASSIC have found that models with a diverse collection of chunks are faster at learning new words and show stronger performance on nonword repetition and sentence recall tasks (Jones & Rowland, 2017). Thus, CIPAL offers a robust chunking mechanism that can discover words in continuous speech and model the behavioral preferences observed in statistical learning experiments without sacrificing these essential nonlexical representations.

As a first approximation, our studies also show that the relationship between lexical processing speed, language experience, and vocabulary size observed in LWL studies (Hurtado et al., 2008; Weisleder & Fernald, 2013) can be explained by a chunking model where reaction times reflect the strength of the individual chunks being used to process the input. By extension, the model also predicts that individual differences in speed of processing emerge from variance in experience with the specific items used in the task (e.g., doggie, baby, ball, shoe). We suggest that children who receive larger and more varied input are more likely to acquire chunks for the items, and these chunks are likely to have faster processing times. This explanation is consistent with correlational studies showing that children who hear greater and more diverse child-directed speech tend to have larger vocabulary sizes (Hart & Risley, 1995; Huttenlocher et al., 1991, 2010; Rowe, 2012) and have faster reaction times in the LWL task across development (Hurtado et al., 2008; Weisleder & Fernald, 2013). However, previous research is divided on whether lexical processing speed reflects a global increase in processing ability or chunk-specific decreases in reaction times (Donnelly & Kidd, 2020). This issue could be addressed within CIPAL in future work by comparing the performance of models that respond to language experience by adjusting the processing times of all chunks (i.e., global capacity) or just the specific chunks used to recode the input, as is assumed here.

Although CIPAL can explain how children identify the sound sequences that correspond to words in their first language, it is not a comprehensive theory of vocabulary development. Brent and Cartwright (1996) suggested that vocabulary acquisition involves learning the phonological form of a word, the semantic concepts it represents, and the syntactic functions it typically performs. They argued that word discovery is solved predominantly by the mechanisms that learn phonological word forms, allowing the semantic and syntactic components to work with the most likely word candidates. This has become the default perspective throughout the word discovery literature. For instance, the artificial languages used in statistical learning experiments typically have no semantic content or syntactic structure. Also, most computational models that operate on child-directed speech corpora, including CIPAL, exploit the distributional structure but not the syntactic or semantic information in the data (e.g., Brent & Cartwright, 1996; Christiansen et al., 1998; French et al., 2011; Goldwater et al., 2009; Monaghan & Christiansen, 2010). Estes et al. (2007) showed that children can discover words embedded in a continuous artificial language and



then use these words as object labels, consistent with the idea that learning the form and the meaning of words are two separate processes that can be studied independently.

However, across different languages, the first words that children can understand and produce are not simply the most frequent patterns, they also tend to have a concrete meaning (Braginsky et al., 2019; Tardif et al., 2008). For instance, children talk about people, animals, and food before they start using function words (e.g., *that*, *how*), which do not have clear semantics but are extremely frequent in their language input. This suggests there is a tight integration between the different components of vocabulary development. We suggest that CIPAL could provide the foundation for a unified theory of vocabulary development that covers the acquisition of phonological word forms, semantics, and morphosyntax (e.g., Gobet et al., 2007). This would impose additional constraints on the model's learning and processing since the different subcomponents would need to work in unison (Byrne, 2012; Gobet & Ritter, 2000; Newell, 1990). By introducing additional mechanisms that are sensitive to semantic information, CIPAL would be able to explore how children connect the meaning of language to the chunks they discover and how this integration shapes their vocabulary acquisition.

## Conclusion

Our goal in this work was to develop a new integrated theory that aligns chunk-based accounts of statistical learning with research in other areas of language development. We introduce the CIPAL architecture, which builds a diverse collection of chunks, including words, multiword phrases, and sublexical units, from patterns in the input. The theory also assumes that these chunks have different representational strengths, where regularly accessed chunks have faster processing times. Through an extensive set of simulation studies, we show that CIPAL can explain word discovery, implicit statistical learning, and speed of lexical processing in different languages and experimental tasks. We argue that incremental chunking is an effective statistical learning mechanism that is central to lexical development, and that children's sensitivity to the statistics of language is a consequence of learning and not the mechanism that drives it.

## References

- Alhama, R. G., & Zuidema, W. (2017). Segmentation as retention and recognition: The R&R model. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual conference of the Cognitive Science Society (CogSci 2017)* (pp. 1531–1536). Cognitive Science Society. [https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item\\_3003208](https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3003208)
- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(2), 239–273. <https://doi.org/10.1017/S030500091400049X>
- Arnon, I. (2021). The Starting Big approach to language learning. *Journal of Child Language*, 48(5), 937–958. <https://doi.org/10.1017/S0305000921000386>
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, 9(3), 621–636. <https://doi.org/10.1111/tops.12271>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82. <https://doi.org/10.1016/j.jml.2009.09.005>
- Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1–2), Article e1373. <https://doi.org/10.1002/wcs.1373>
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324. <https://doi.org/10.1111/1467-9280.00063>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- 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(3), 241–248. <https://doi.org/10.1111/j.1467-9280.2008.02075.x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bartha-Doering, L., Deuster, D., Giordano, V., am Zehnhoff-Dinnesen, A., & Döbel, C. (2015). A systematic review of the mismatch negativity as an index for auditory sensory memory: From basic research to clinical and developmental perspectives. *Psychophysiology*, 52(9), 1115–1130. <https://doi.org/10.1111/psyp.12459>
- Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83(2), 167–206. [https://doi.org/10.1016/S0010-0277\(02\)00002-1](https://doi.org/10.1016/S0010-0277(02)00002-1)
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. <https://doi.org/10.48550/arXiv.1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Bergelson, E., & Swingle, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, 11(4), 369–380. <https://doi.org/10.1080/15475441.2014.979387>
- Bergelson, E., & Swingle, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, 89(5), 1567–1576. <https://doi.org/10.1111/cdev.12888>
- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual meeting of the Cognitive Science Society* (pp. 124–129). Cognitive Science Society. <https://hdl.handle.net/11858/00-001M-0000-002D-DEDE-0>
- Blanchard, D., Heinz, J., & Golinkoff, R. (2010). Modeling the contribution of phonotactic cues to the problem of word segmentation. *Journal of Child Language*, 37(3), 487–511. <https://doi.org/10.1017/S030500090999050X>
- Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development*, 87(6), 1893–1908. <https://doi.org/10.1111/cdev.12554>
- Braginsky, M., Sanchez, A., & Yurovsky, D. (2022). *childesr: Accessing the "CHILDES" database* (Version 0.2.3) [Computer software]. <https://CRA-N.R-project.org/package=childesr>
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, 3, 52–67. [https://doi.org/10.1162/opmi\\_a\\_00026](https://doi.org/10.1162/opmi_a_00026)



- Brent, M. R. (1999). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 34(1), 71–105. <https://doi.org/10.1023/A:1007541817488>
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1–2), 93–125. [https://doi.org/10.1016/S0010-0277\(96\)00719-6](https://doi.org/10.1016/S0010-0277(96)00719-6)
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733. <https://doi.org/10.1353/lan.2006.0186>
- Byrne, M. D. (2012). Unified theories of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4), 431–438. <https://doi.org/10.1002/wcs.1180>
- Cabiddu, F., Bott, L., Jones, G., & Gambi, C. (2023). CLASSIC utterance boundary: A chunking-based model of early naturalistic word segmentation. *Language Learning*, 73(3), 942–975. <https://doi.org/10.1111/la.ng.12559>
- Caines, A., Altmann-Richer, E., & Buttery, P. (2019). The cross-linguistic performance of word segmentation models over time. *Journal of Child Language*, 46(6), 1169–1201. <https://doi.org/10.1017/S0305000919000485>
- Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33(2), 111–153. <https://doi.org/10.1006/cogp.1997.0649>
- Cheour, M., Ceponienė, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V., & Näätänen, R. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology*, 43(1), 33–39. <https://doi.org/10.1111/1467-9450.00266>
- Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, 11(3), 468–481. <https://doi.org/10.1111/tops.12332>
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13(2–3), 221–268. <https://doi.org/10.1080/016909698386528>
- Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, 9(3), 542–551. <https://doi.org/10.1111/tops.12274>
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Lawrence Erlbaum.
- Contreras Kallens, P., & Christiansen, M. H. (2022). Models of language and multiword expressions. *Frontiers in Artificial Intelligence*, 5, Article 781962. <https://doi.org/10.3389/frai.2022.781962>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, 128(2), 165–185. <https://doi.org/10.1037/0096-3445.128.2.165>
- de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Van Gorcum.
- Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, 9(1), 1–7. <https://doi.org/10.1111/1467-9280.00001>
- Donnelly, S., & Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation. *Journal of Experimental Child Psychology*, 192, Article 104781. <https://doi.org/10.1016/j.jecp.2019.104781>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. [https://doi.org/10.1207/s15516709cog1402\\_1](https://doi.org/10.1207/s15516709cog1402_1)
- Endress, A. D., & Johnson, S. P. (2021). When forgetting fosters learning: A neural network model for statistical learning. *Cognition*, 213, Article 104621. <https://doi.org/10.1016/j.cognition.2021.104621>
- Endress, A. D., & Langus, A. (2017). Transitional probabilities count more than frequency, but might not be used for memorization. *Cognitive Psychology*, 92, 37–64. <https://doi.org/10.1016/j.cogpsych.2016.11.004>
- Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, 60(3), 351–367. <https://doi.org/10.1016/j.jml.2008.10.003>
- Endress, A. D., Slone, L. K., & Johnson, S. P. (2020). Statistical learning and memory. *Cognition*, 204, Article 104346. <https://doi.org/10.1016/j.cognition.2020.104346>
- eSpeak NG Text-to-Speech (Version 1.51) [Computer software] (2022). <https://github.com/espeak-ng/espeak-ng>
- Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science*, 18(3), 254–260. <https://doi.org/10.1111/j.1467-9280.2007.01885.x>
- Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science*, 8(4), 305–336. [https://doi.org/10.1207/s15516709cog0804\\_1](https://doi.org/10.1207/s15516709cog0804_1)
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), i–185. <https://doi.org/10.2307/1166093>
- 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. <https://doi.org/10.1037/0012-1649.42.1.98>
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231. <https://doi.org/10.1111/1467-9280.00044>
- Fernald, A., Borschinger, B., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: Online methods in children's language processing* (pp. 97–135). John Benjamins Publishing. <https://doi.org/10.1075/lald.44.06fer>
- Fourtassi, A., Borschinger, B., Johnson, M., & Dupoux, E. (2013). Why is English so easy to segment? *Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics* (pp. 1–10). Association for Computational Linguistics. <https://aclanthology.org/W13-2601/>
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. <https://doi.org/10.1017/S0305000916000209>
- Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, 117(2), 107–125. <https://doi.org/10.1016/j.cognition.2010.07.005>
- French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation

- and chunk extraction. *Psychological Review*, 118(4), 614–636. <https://doi.org/10.1037/a0025255>
- Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, 31(2), 311–341. <https://doi.org/10.1080/15326900701221454>
- Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, 145(12), 1128–1153. <https://doi.org/10.1037/bul0000210>
- 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(1), 83–94. <https://doi.org/10.3758/BF03210559>
- Gervain, J., & Guevara Erra, R. (2012). The statistical signature of morphosyntax: A study of Hungarian and Italian infant-directed speech. *Cognition*, 125(2), 263–287. <https://doi.org/10.1016/j.cognition.2012.06.010>
- Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, 33(2), 260–272. <https://doi.org/10.1111/j.1551-6709.2009.01012.x>
- Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition*, 66(2), 115–152. [https://doi.org/10.1016/S0010-0277\(98\)00020-1](https://doi.org/10.1016/S0010-0277(98)00020-1)
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243. [https://doi.org/10.1016/S1364-6613\(00\)01662-4](https://doi.org/10.1016/S1364-6613(00)01662-4)
- Gobet, F., Pine, J. M., & Freudenthal, D. (2007). Towards a unified model of language acquisition. *Proceedings of the European Cognitive Science Conference* (pp. 602–607). European Cognitive Science Society. <https://bura.brunel.ac.uk/handle/2438/694>
- Gobet, F., & Ritter, F. E. (2000). Individual data analysis and unified theories of cognition: A methodological proposal. *Proceedings of the 3rd International Conference on Cognitive Modelling* (pp. 150–157). Universal Press.
- Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, 112(1), 21–54. <https://doi.org/10.1016/j.cognition.2009.03.008>
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(3), 515–531. <https://doi.org/10.1017/S0305000907008641>
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31(2), 190–222. <https://doi.org/10.2307/411036>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2), 185–207. <https://doi.org/10.3758/BF03212979>
- Huetttig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. <https://doi.org/10.1016/j.actpsy.2010.11.003>
- Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, 17(1), 189–202. <https://doi.org/10.1017/S136672891300014X>
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), F31–F39. <https://doi.org/10.1111/j.1467-7687.2008.00768.x>
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236–248. <https://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(4), 343–365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Isbilen, E. S., & Christiansen, M. H. (2020). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*, 12(2), 713–726. <https://doi.org/10.1111/tops.12376>
- Jarecki, J. B., Tan, J. H., & Jenny, M. A. (2020). A framework for building cognitive process models. *Psychonomic Bulletin & Review*, 27(6), 1218–1229. <https://doi.org/10.3758/s13423-020-01747-2>
- Jarosz, G., & Johnson, J. A. (2013). The richness of distributional cues to word boundaries in speech to young children. *Language Learning and Development*, 9(2), 175–210. <https://doi.org/10.1080/15475441.2011.641904>
- 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. <https://doi.org/10.1111/desc.12111>
- Jones, G., Gobet, F., & Pine, J. M. (2005). Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory. *Artificial Intelligence and Simulation of Behaviour Journal*, 1(6), 509–522. <https://irep.ntu.ac.uk/id/eprint/12034>
- 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. <https://doi.org/10.1111/j.1467-7687.2007.00638.x>
- 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. <https://doi.org/10.1016/j.cogpsych.2017.07.002>
- Junge, C. (2017). The proto-lexicon: Segmenting word-like units from the speech stream. In G. Westermann & N. Mani (Eds.), *Early word learning* (pp. 15–29). Routledge. <https://doi.org/10.4324/9781315730974-2>
- Juszyk, P. W., & Aslin, R. N. (1995). Infants’ detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23. <https://doi.org/10.1006/cogp.1995.1010>
- Ko, E. S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua*, 122(8), 841–857. <https://doi.org/10.1016/j.lingua.2012.02.005>
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843. <https://doi.org/10.1038/nrn1533>
- Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, 127(3), 439–453. <https://doi.org/10.1016/j.cognition.2013.02.002>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lane, P. C. R., & Gobet, F. (2003). Developing reproducible and comprehensible computational models. *Artificial Intelligence*, 144(1–2), 251–263. [https://doi.org/10.1016/S0004-3702\(02\)00384-3](https://doi.org/10.1016/S0004-3702(02)00384-3)
- Lane, P. C. R., & Gobet, F. (2012). A theory-driven testing methodology for developing scientific software. *Journal of Experimental & Theoretical Artificial Intelligence*, 24(4), 421–456. <https://doi.org/10.1080/0952813X.2012.695443>
- Lany, J., Giglio, M., & Oswald, M. (2018). Infants’ lexical processing efficiency is related to vocabulary size by one year of age. *Infancy*, 23(3), 342–366. <https://doi.org/10.1111/inf.12228>
- Li, C. (2019). JuliaCall: An R package for seamless integration between R and Julia. *Journal of Open Source Software*, 4(35), Article 1284. <https://doi.org/10.21105/joss.01284>
- Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120(11), 2546–2556. <https://doi.org/10.1016/j.lingua.2010.06.005>

- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 883–914. <https://doi.org/10.1037/0278-7393.18.5.883>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
- MacWhinney, B. (2000). *The childes project: Tools for analyzing talk: Vol. II. The database* (3rd ed.). Lawrence Erlbaum. <https://doi.org/10.4324/9781315805641>
- 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. <https://doi.org/10.1017/S0305000909990055>
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500. <https://doi.org/10.1037/0096-3445.134.4.477>
- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, 126(1), 1–51. <https://doi.org/10.1037/rev0000126>
- McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631. <https://doi.org/10.1126/science.1144073>
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41(4), 215–225. <https://doi.org/10.1159/000261728>
- Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in english-learning 15-month-olds. *Frontiers in Psychology*, 4, Article 24. <https://doi.org/10.3389/fpsyg.2013.00024>
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315373218>
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. <https://doi.org/10.1016/j.jml.2007.11.006>
- Mirman, D., Graf Estes, K., & Magnuson, J. S. (2010). Computational modeling of statistical learning: Effects of transitional probability versus frequency and links to word learning. *Infancy*, 15(5), 471–486. <https://doi.org/10.1111/j.1532-7078.2009.00023.x>
- Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, 37(3), 545–564. <https://doi.org/10.1017/S0305000909990511>
- Näätänen, R. (1992). *Attention and brain function*. Psychology Press.
- Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, 139(3), 1272–1281. <https://doi.org/10.1121/1.4944634>
- Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Psychology Press.
- Norris, D., & Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition*, 208, Article 104534. <https://doi.org/10.1016/j.cognition.2020.104534>
- Ordin, M., Polyanskaya, L., & Soto, D. (2020). Neural bases of learning and recognition of statistical regularities. *Annals of the New York Academy of Sciences*, 1467(1), 60–76. <https://doi.org/10.1111/nyas.14299>
- Ordin, M., Polyanskaya, L., Soto, D., & Molinaro, N. (2020). Electro-physiology of statistical learning: Exploring the online learning process and offline learning product. *European Journal of Neuroscience*, 51(9), 2008–2022. <https://doi.org/10.1111/ejn.14657>
- Pearl, L., Goldwater, S., & Steyvers, M. (2010). Online learning mechanisms for Bayesian models of word segmentation. *Research on Language and Computation*, 8(2), 107–132. <https://doi.org/10.1007/s11168-011-9074-5>
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244–247. <https://doi.org/10.1016/j.cognition.2009.07.011>
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674–685. <https://doi.org/10.1111/j.1467-8624.2009.01290.x>
- Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, 11(3), 520–535. <https://doi.org/10.1111/tops.12403>
- Perruchet, P., & Desauty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, 36(7), 1299–1305. <https://doi.org/10.3758/MC.36.7.1299>
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233–238. <https://doi.org/10.1016/j.tics.2006.03.006>
- Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, 66(4), 807–818. <https://doi.org/10.1016/j.jml.2012.02.010>
- Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*, 34(2), 255–285. <https://doi.org/10.1111/j.1551-6709.2009.01074.x>
- Perruchet, P., & Vinter, A. (1998). PARSE: A model for word segmentation. *Journal of Memory and Language*, 39(2), 246–263. <https://doi.org/10.1006/jmla.1998.2576>
- 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, Article 101238. <https://doi.org/10.1016/j.cogpsych.2019.101238>
- Phillips, L., & Pearl, L. (2014). Bayesian inference as a cross-linguistic word segmentation strategy: Always learning useful things. *Proceedings of the 5th Workshop on Cognitive Aspects of Computational Language Learning (CogACL)* (pp. 9–13). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-0503>
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130. <https://doi.org/10.3758/s13423-014-0585-6>
- Polyanskaya, L. (2022). Cognitive mechanisms of statistical learning and segmentation of continuous sensory input. *Memory & Cognition*, 50(5), 979–996. <https://doi.org/10.3758/s13421-021-01264-0>
- Poulin-Charronnat, B., Perruchet, P., Tillmann, B., & Peereman, R. (2017). Familiar units prevail over statistical cues in word segmentation. *Psychological Research*, 81(5), 990–1003. <https://doi.org/10.1007/s00426-016-0793-y>
- Raneri, D., Von Holzen, K., Newman, R., & Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, 47(6), 1263–1275. <https://doi.org/10.1017/S030500091900093X>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Real, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, 60(2), 161–170. <https://doi.org/10.1080/17470210600971469>



- Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: A MDL-based cognitive model of inductive learning. *Cognitive Science*, 35(7), 1352–1389. <https://doi.org/10.1111/j.1551-6709.2011.01188.x>
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5), 1762–1774. <https://doi.org/10.1111/j.1467-8624.2012.01805.x>
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81(2), 149–169. [https://doi.org/10.1016/S0010-0277\(01\)00132-9](https://doi.org/10.1016/S0010-0277(01)00132-9)
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, 69(1), 181–203. <https://doi.org/10.1146/annurev-psych-122216-011805>
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. <https://doi.org/10.1006/jmla.1996.0032>
- Saksida, A., Langus, A., & Nespor, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science*, 20(3), Article e12390. <https://doi.org/10.1111/desc.12390>
- Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 592–608. <https://doi.org/10.1037/0278-7393.16.4.592>
- Slone, L. K., & Johnson, S. P. (2015). Statistical and chunking processes in adults' visual sequence learning. *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2218–2223). Cognitive Science Society.
- Slone, L. K., & Johnson, S. P. (2018). When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition*, 178, 92–102. <https://doi.org/10.1016/j.cognition.2018.05.016>
- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4), 869–902. <https://doi.org/10.1017/S0305000908008763>
- Swingle, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50(1), 86–132. <https://doi.org/10.1016/j.cogpsych.2004.06.001>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. <https://doi.org/10.1126/science.7777863>
- Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, 44(4), 929–938. <https://doi.org/10.1037/0012-1649.44.4.929>
- Theakston, A., & Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science*, 9(3), 588–603. <https://doi.org/10.1111/tops.12268>
- Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), Article 20160056. <https://doi.org/10.1098/rstb.2016.0056>
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39(4), 706–716. <https://doi.org/10.1037/0012-1649.39.4.706>
- Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3(1), 73–100. <https://doi.org/10.1080/15475440709337001>
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- Venkataraman, A. (2001). A statistical model for word discovery in transcribed speech. *Computational Linguistics*, 27(3), 351–372. <https://doi.org/10.1162/089120101317066113>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>

Received July 4, 2023

Revision received March 17, 2025

Accepted March 24, 2025 ■