Insights from studying statistical learning

Rebecca L. A. Frost and Padraic Monaghan
Max Planck Institute for Psycholinguistics / University of Lancaster

Acquiring language is notoriously complex, yet for the majority of children this feat is accomplished with remarkable ease. Usage-based accounts of language acquisition suggest that this success can be largely attributed to the wealth of experience with language that children accumulate over the course of language acquisition. One field of research that is heavily underpinned by this principle of experience is statistical learning, which posits that learners can perform powerful computations over the distribution of information in a given input, which can help them to discern precisely how that input is structured, and how it operates. A growing body of work brings this notion to bear in the field of language acquisition, due to a developing understanding of the richness of the statistical information contained in speech. In this chapter we discuss the role that statistical learning plays in language acquisition, emphasising the importance of both the distribution of information within language, and the situation in which language is being learnt. First, we address the types of statistical learning that apply to a range of language learning tasks, asking whether the statistical processes purported to support language learning are the same or distinct across different tasks in language acquisition. Second, we expand the perspective on what counts as environmental input, by determining how statistical learning operates over the situated learning environment, and not just sequences of sounds in utterances. Finally, we address the role of variability in children's input, and examine how statistical learning can accommodate (and perhaps even exploit) this during language acquisition.

Preface

Elena Lieven has provided some of the best evidence for how children's language development is directly dependent upon their language experience. The empiricist, usage-based principles that characterise her approach have been enormously influential in terms of defining both the types of processes and mechanisms that are required to explain children's behaviour, and the language knowledge that

https://doi.org/10.1075/tilar.27.03fro

underlies that behaviour. In conversations with Elena over the last few years, in regular reading group meetings in North West England, and then through the thrilling collaborations that have emerged during the ESRC International Centre for Language and Communicative Development, the similarities between this perspective and our statistical learning approach have (at least on our side of these conversations) been strikingly apparent. Nevertheless, Elena's scepticism about the artificial nature of our statistical language learning studies has proven a valuable fillip to ensure that our understanding of language is properly respected, in terms of key characteristics of its structure, and the conditions within which language learning is embedded during the natural interlocutory experience of the child (see also Dupoux, 2018). In this chapter, we review some of our reflections on statistical learning in language acquisition, which owe a debt to Elena's approach that emphasises the importance of not just looking to the distribution, but also considering the language learning situation in all its florid, glorious, and sometimes puzzling, complexity.

Introduction

Within a few short years, children develop a degree of knowledge about language close to mastery, sufficient to comprehend fast, varied, incoming speech, and sufficient to use language productively. Understanding humans' path to linguistic proficiency has been a mainstay of language philosophy and language acquisition research since the earliest recorded debates between nativist and empiricist accounts of learning.

Arguments over whether (language) knowledge is innate or acquired have resounded throughout the history of thought. In *The Republic*, Plato stated the case for knowledge to be innate because the form of the knowledge to be acquired is not perfectly contained in the senses (thus, the senses require perceptions in order to interpret the abstract form that generates the percept itself). In counterpoint, in *The Posterior Analytics*, Aristotle questioned the possible existence of innate knowledge prior to experience, concluding instead that knowledge was a consequence of extracting generalisations from repeated experience (Modrak, 2001). Such contrasting views have repeated in the 20th and 21st Centuries with particular reference to the domain of linguistic structure – with nativist accounts theorising that language structure is innately encoded knowledge that is activated through recognition of percepts, while empiricist views postulate that it can be entirely extracted from general purpose cognitive operations over experience. In the empiricist tradition, frequently repeating language structures were seen to provide the possibility for structure learning. For instance, Harris (1954) proposed that the meaning of a word, and its usage, could be determined from co-occurrences of that word within the language. For grammatical categories, Fries (1952) demonstrated that co-occurrences between very frequent words and the words that precede or follow them could fall into distinct clusters with little overlap, which could be described in terms of the grammatical categories to which those words belong. For instance, nouns frequently follow articles (*the*, *a*) but verbs seldom do, whereas nouns seldom follow pronouns (*you*, *I*), but verbs often do.

However, these early demonstrations of the potential for experience to define language structures were beset by the problem of very small corpora, often in a single language (Fries, 1952; Harris, 1954), thus a proof of concept was not sufficient proof that adequate learning was possible from experience alone. Furthermore, accounts concerning the range of cognitive computational mechanisms that could operate over language input were somewhat limited, leaving suggestions for the possible limits of such *statistical learning* overly constrained (Redington et al., 1998). For instance, Pinker (1984) argued that, "the properties that the child can detect in the input – such as the serial positions and adjacency and cooccurrence relations among words – are in general linguistically irrelevant." Relatedly, there remained the continued puzzle as to how structures that had never occurred during exposure could be acquired if learning was based entirely on experience (Chomsky, 1981; Crain & Nakayama, 1987). Thus, proposals for nativist, generativist accounts of language learning became the dominant descriptions of language acquisition throughout the 20th Century (see review by Pullum & Scholz, 2002).

Against this mainstream nativist view, work on the role of experience for language learning continued. Prominent amongst these approaches were the observations by Lieven and colleagues that children's language productions could be described in terms of reproduction and manipulation of linguistic constructions that the individual child is exposed to through the productions of their caregivers. Lieven and colleagues thus sought to determine precisely how experience gives rise to knowledge of language structure (e.g., Lieven et al., 2003, and see also Cameron-Faulkner et al., 2003; Lieven & Brandt, 2011; Lieven et al., 2009). This important work helped to define our understanding of the complexity of the computational processes that are available to the learner, and the consequent richness of the structure that can be induced through the operation of these processes over the child's input. Such a view enables us to determine the extent to which language structure can be learned from statistical computational processes. Furthermore, it permits theorising on whether the computational processes children perform during language acquisition employ language-specific mechanisms that assist language acquisition alone, or general purpose mechanisms that help learning more broadly and are brought to bear to help with the special task of acquiring language.

Drawing on many of the same key principles as Lieven and colleagues' work on experience-based learning, the field of statistical learning has revealed substantial insights into the way in which children can learn from the linguistic input they receive – particularly with regard to acquisition of language structure. Research on statistical language learning seeks to uncover the precise distributional properties of children's input, and determine the operations that children can perform over that input that may account for their developing (implicit) knowledge of the language, sufficient to explain their growing linguistic productivity. The assumptions underpinning this literature counter alternative accounts that propose language structure is not acquired via computations over the distributional properties of the input, but rather that the input activates specific innate knowledge about language. Under the empiricist approach, it is assumed (until observation proves this approach inadequate) that children learn about language from experience, and respond to the linguistic properties that are sufficiently expressed within this language experience itself.

In this chapter, we consider two key themes regarding progress in studies of statistical learning of language. First, we address the types of statistical learning that apply to a range of language learning tasks, asking whether the statistical processes purported to support language learning are the same or distinct across different tasks in language acquisition. Second, we expand the perspective on environmental input, by determining how statistical learning operates over the whole environment, and not just sequences of sounds in utterances, to address the question of how situated language affects learning. A key aspect of language learning is the variability that permeates the learning environment, and so we also address the role of variability in children's input, and examine how statistical learning accommodates – and even exploits – this variability.

Statistical processes for different language learning tasks

Contained within language is a panoply of information on which learners can draw to help them master a broad range of language learning tasks. At every level of linguistic structure exist regularities that can assist learning – for instance, co-occurrence of particular phones can help learners to develop knowledge of the phonotactics within syllables, while co-occurrences of those syllables can come to constrain how words are constructed in the language. Similarly, information about how those words are used in combination helps learners to discern how words group together in utterances, and how the language operates in terms of its grammatical and syntactic structure. Investigating learners' ability to extract and compute over language input to derive these levels of linguistic structure is critical for our understanding of language learning. Whether and how the nature of this learning differs across tasks, though, is subject to debate – particularly with regard to early stages of acquisition.

For speech segmentation, the role of a distributional learning mechanism has been long accepted, with accounts largely converging on the notion that statistical processing drives acquisition of the ability to segment speech into words. There has been substantial progress in defining the statistical mechanism that children apply to speech input in order to master this skill, stemming from the seminal study by Saffran, Aslin, & Newport (1996), which discovered that children were sensitive to transitional probabilities between syllables in speech. Detecting and computing over transitional probability variation can assist speech segmentation since transitional probabilities between words are low, relative to transitional probabilities between syllables within words, which are often much higher. For instance, in the CHILDES database (MacWhinney, 2000), the transitional probability between "the" and "peng" is .1% – among 5.5M words of English child-directed speech there are 179 instances of this transition, and 149,743 instances where "the" is succeeded by something other than "peng". In contrast, "peng" is succeeded by "uin" 601 out of 609 times in the same corpus, yielding a transitional probability of 98.7%.

Saffran and colleagues (1996) demonstrated that 8-month-old infants could track co-occurrences between adjacent syllables in speech, and distinguish between syllable transitions that had different probabilities of co-occurrence. The study involved highly constrained artificial speech that was produced in a continuous stream, comprising four trisyllabic words. Particular triplets of syllables always occurred together, forming the words, but each word could be followed by any one of the other three words. Thus, within words transitional probabilities between syllables were 1, whereas between words transitional probabilities were much lower (0.33). After listening to the speech, children were tested on their looking times to stimuli that comprised words in the language, and stimuli that comprised trisyllabic strings that spanned two words (*part-words*). If learners were sensitive to the distributional properties of the language they had heard, and if they could compute over these properties in such a way that informed learning, then they may respond differently to the words and part-words at test. Indeed, during testing, children looked for longer at the part-words than the words, indicating a novelty preference for the sequences that contained lower transitional probabilities – implicating a possible role for statistical learning in speech segmentation.

This statistical segmentation effect has been replicated many times in subsequent research, for both infant and adult learners (e.g., Saffran et al., 1997), and is corroborated in a recent meta-analysis by Black and Bergmann (2017). Perhaps even more convincingly, evidence for statistical segmentation has also been found for natural (as well as artificial) language stimuli (Pelucchi et al., 2009), lending great credence to the notion that a mechanism capable of computing across syllable transitions in speech may contribute to children's language acquisition.

Much of the work that grounds our understanding of statistical speech segmentation has focussed on computing over syllables that follow one another in speech directly. However, transitional probabilities between non-adjacent syllables (i.e., syllables that are separated in speech, but are critically dependent on one another) have also been shown to help learners segment speech into candidate words (Marchetto & Bonatti, 2015; Peña et al., 2002). In these studies, words comprised syllable triplets of the form AXC, with each letter indicating a different syllable. Unlike the Saffran et al. stimuli, here there was a transitional probability of 1 between the first and third syllable within a word (syllables A and C), whereas the transition between the first and second (A-X), and second and third syllables (X-C) had a probability that was much lower (0.33 in Peña et al.'s study, and 0.5 in Marchetto and Bonatti's). In Peña et al.'s study with three nonadjacent pairings, the transitional probability between the first and second, and second and third syllables was therefore the same as that between the final syllable of the word and the beginning of the next word – meaning learners had to look to the nonadjacent dependencies to help extract the words from speech (though these statistics differ slightly in infant studies, as the language-size is typically reduced to two, rather than three, non-adjacent dependencies).

After hearing continuous speech comprising nonadjacent dependencies, infants over 12 months of age have been shown to distinguish between words and non-word competitors, demonstrated through longer listening to non-words during testing. Similarly, adult learners have been shown to select words over part-words as more likely word candidates on a two-alternative forced-choice test. These findings have been replicated for both infant (Frost et al., 2020) and adult learners (e.g., Frost & Monaghan, 2016), providing converging evidence that learners can draw on non-adjacent distributional relationships to help them to segment speech – even though this requires computing over intervening material. In terms of generalising the within word-structures, learner's statistical learning abilities are somewhat different: adults can generalise these dependencies to novel instances, even when trained on continuous speech – meaning they can segment the words from speech, and learn to generalise the within-word structure at the same time, from statistical information alone (Frost & Monaghan, 2016). For infants though, there is little evidence that these tasks can proceed in tandem from statistics alone (e.g., Marchetto & Bonatti, 2013, 2015, but see Frost et al., 2020), and research has suggested that these tasks could follow distinct developmental trajectories (Marchetto & Bonatti, 2013, 2015), even though they may draw on similar statistical mechanisms (see Frost & Monaghan, 2016, for further discussion).

Further evidence for the role of transitional probabilities in locating word boundaries can be found in the computational modelling literature. Elman (1993) showed that a recurrent neural network model was sensitive to changes in transitional probabilities when presented with a corpus of continuous speech input, with the model being substantially better at predicting the next item of speech within words, compared to at word boundaries. French et al. (2011) extended this model to show that both forward and backward transitional probabilities could drive learning. This application of sequential statistical learning showed how transitional probabilities are able to highlight word boundaries in continuous speech, though it did not explain how those words might actually be acquired into the learner's vocabulary.

A computational model termed PARSER (Perruchet & Vinter, 1998) took a somewhat different approach to resolving speech segmentation; rather than processing transitional probabilities from speech, Perruchet and Vinter (1998) showed that PARSER learned to join together, or chunk, frequently occurring sequences in speech in order to extract words from artificial language input (such as that used in the Saffran et al. (1996) experiments). PARSER could also demonstrate segmentation for languages containing non-adjacent dependencies (again through chunking, rather than transitional probability computation; Perruchet et al., 2004), simulating the behavioural results observed by Peña et al. (2002), but through critically different means. Thus, demonstrations of learning from input containing variation in transitional probabilities do not necessarily indicate that learning is based on those transitional probabilities. Indeed, Perruchet and Vinter (1998) highlighted that associations between frequent-occurring sequences could sufficiently explain the behavioural results observed in Saffran et al.'s (1996) study of speech segmentation, and Pena et al.'s (2002) study of nonadjacent dependency learning.

Models such as these have substantially enriched our understanding of language acquisition. However, a model that already isolates syllables in effect solves a great deal of the problem of segmentation, because identifying where syllable boundaries fall is a profoundly difficult challenge in its own right. Thus, models that join together frequently occurring phoneme pairs according to their co-occurrence statistics provide a further incremental step toward understanding how learners build up and distinguish words in continuous speech (e.g., Hockema, 2006; Swingley, 2005). The task of identifying phonemes from continuous speech, though, is in itself extremely challenging (Kamper, Jansen, & Goldwater, 2016).

A general feature of these models – both those that rely on transitional probabilities, and those that rely on associations between syllables – is that they build words up by chunking together co-occurring sequences from speech. However, an alternative statistical mechanism that operates on speech input could instead begin with larger sequences (utterances or utterance fragments) and break these down into their potential constituent words. Monaghan and Christiansen (2010) examined the efficacy of such a mechanism with their PUDDLE model of speech segmentation, which they applied to natural language corpora of child-directed speech. The model began with a whole utterance represented as a string of phonemes (rather than syllables), treating the entire string as a candidate word in the language. When a word that already occurred in the language was present in a later utterance, the utterance was broken down into the part preceding the known sequence, and the part succeeding it. In this way, the model learned effectively to isolate words in speech, with words represented in terms of their phonemes rather than their syllables.

Sensitivity to statistical information has been also proposed to support acquisition of morphology. General purpose information-theoretic principles have been suggested to be capable of identifying morphemes (Harris, 1955), as well as word boundaries, and this was taken to be entirely compatible with early generative accounts of how those identified morphemes would then interoperate (Chomsky, 2005). Using the MOSAIC model, Freudenthal, Pine, and colleagues demonstrated how representation of morphological structure can be effectively deduced from the statistical information present in child-directed speech (e.g., Freudenthal et al., 2006). MOSAIC incorporates two key components: a memory store, and statistical mechanisms that parse speech in relation to the stored representations of speech in memory, in order to detect morphemes. The model receives child-directed speech utterances as input, and stores fragments of those utterances, constrained by working memory limitations. Those fragments thus comprise the end of the utterance (as this is the most recently experienced part of the speech within working memory). The proportion of the utterance that can be processed from the end is a property of the number of previously stored fragments that it comprises. Stored fragments gradually increase in size due to recognition of the fragments they contain – enabling larger sequences to be stored in memory. Thus, the model stores subsequences of utterances of varying size. For production, the model retrieves these subsequences, therefore speech comprehension is driven initially by splitting the input, according to memory limitations, and speech production then involves chunking these fragments together. The model is able to simulate children's productive language development, demonstrating key phenomena, such as use of the optional infinitive (e.g., "He go" instead of "He goes") in early language production.

Approaches to word and morphological segmentation have therefore yielded rich results, through a combination of corpus analyses that determine the multiple sources of statistical information in language that could assist learning, computational models that explicitly test the usefulness of this information, and experimental studies that measure learners' sensitivity to this information in speech.

Similar approaches have been taken to determine the statistical mechanisms available for learning mappings between words and their potential referents. A key challenge in language acquisition is determining which referent in the child's environment a word refers to, from a potentially infinite array of possibilities (Quine, 1960). Proposals for how children constrain possible mappings have focused either on innate or strategic constraints applying to the communicative situation (such as mutual exclusivity – assuming a novel word maps to an unknown object, or the whole-object constraint – assuming a label refers to an entire object rather than one of the parts comprising it; see Monaghan et al., 2017, for a review). Alternative accounts have explored the extent to which learning can instead be driven by the development of statistical associations between particular words and features of the environment, based on the occurrences (and co-occurrences) of these words and features over multiple learning instances.

Smith and Yu (2008) showed that infants were able to track such cross-situational statistics to acquire mappings between particular words and objects, and research using head-mounted eyetracking is beginning to explore the way that the dynamic statistical properties of children's environments shape their language acquisition in more naturalistic settings than the laboratory (Clerkin et al., 2017; Smith et al., 2018). Children's capacity to learn words through computing cross-situational statistics has not only been shown for learning mappings between words and objects (i.e., nouns); in similar laboratory-based research, Scott and Fisher (2012) showed that such cross-situational statistics can equally apply to acquisition of word–action mapping in learning verb referents, and Monaghan et al. (2015) showed that both nouns and verbs could be acquired simultaneously from cross-situational co-occurrence statistics. Computational models of word-referent learning have explored whether associative information alone is sufficient for learning new mappings, or whether referent selection strategies (such as mutual exclusivity) are also important during word learning tasks. McMurray et al. (2012) proposed that both mechanisms – fast mapping for referent selection, and slow associative learning for gradual accumulation of word–referent mappings – accounted for performance, whereas Yu and Smith (2012) showed instead that associative statistics alone were sufficient to reflect children's word learning behaviour.

In addition to identifying words from speech, and linking them to referents in their environment, learners must also develop an understanding of the role of those words in utterances. The extent to which the grammatical category of a word can be acquired from statistical information has remained a question of debate in language acquisition research. Since Fries' (1952) initial small-scale studies of relations between distributions of words and the grammatical categories to which those words belong, studies have assessed the availability of distributional information in larger, more realistic corpora of child-directed speech. Redington et al. (1998) demonstrated that statistical co-occurrences were sufficient to provide grammatical categories for words to a high degree of accuracy, and in addition provided nuanced information about the blurred borders between grammatical categories reflecting their variable usage in child-directed speech (e.g., some words can be used as both nouns and verbs). Mintz (2003) showed that not only was distributional statistical information useful for discovering grammatical categories of words, but that this information was also tractable, obviating the need for learners to process and store very many associations between words – as was the case in Redington et al.'s (1998) corpus study. Mintz (2003) found that a small number of frequent frames (frequently co-occurring words separated in speech by one other word) could provide accurate information about grammatical categories of words that occur inside the frame. For instance, nouns can occur within the frame "the __ is" but verbs cannot, and verbs can occur within "he ___ the" but nouns cannot. St Clair et al. (2010) showed that these frequent frames were somewhat sparse in terms of their coverage of the words in the child's early vocabulary, but that a combination of simpler frames "he ___" with "___ the" could provide similarly accurate categorisations of a far larger proportion of the child's language exposure than in Mintz' (2003) analysis. Such flexible frames are also more likely to apply cross-linguistically (see, e.g., Stumper et al., 2011).

Thus, the referent for a word and the grammatical category to which that word belongs are potentially derivable from cross-situational statistics that track co-occurrences between objects/actions and particular words. Further, additional distributional information within utterances can constrain learners' hypotheses about which words are likely referents for objects and which are words from other grammatical categories (Monaghan & Mattock, 2012). Nativist accounts of relations between words and grammatical categories (concerning innate semantic features) are thus shown to be unnecessary for acquiring both a word's meaning and the grammatical category to which that word belongs.

Statistical learning has been found capable of accounting for a multitude of tasks in language learning with substantial success, from speech segmentation to grammatical categorisation. Thus, it remains a curiosity as to why acquisition of syntax has traditionally been considered an exception to the rule of statistical learning mechanisms. However, there are long-standing traditions within syntax acquisition research that take precisely this approach. For instance, within the usage-based framework Lieven and colleagues have shown that the hallmark of processing for syntax acquisition is to acquire sequences of words from caregivers' productions and manipulate them in a productive language system. In Lieven et al. (2003) the utterances of a 2-year-old child were analysed for the extent to which they reproduced utterance constructions produced by their caregivers, with the authors specifying what operations could give rise to the child's productions if they were derivations, but not precise reproductions, of their parent's utterances.

Lieven et al. (2003) proposed five possible manipulations by which constructions of the parent's speech could generate novel constructions produced by the child: The child could substitute one word within a phrase (e.g., I want a paper -> I want a penguin), the child could add two constructions together (let's move it -> let's move it around), drop a part from the beginning or the end of a construction (have you finished your book -> have you finished), remove or insert a previously occurring word into the current construction (have you finished your book -> have you finished with your book), or rearrange words within a construction (e.g., away it goes -> it goes away). Of all the instances where the child's production was related to their parent's utterance by one or more of these manipulations, 66% were substitutions, 15% were additions, 10% were removals, 7% were insertions, and 1% were rearrangements. These construction manipulations, particularly adding and dropping, indicate that both chunking and dividing mechanisms are critically involved in language acquisition. Deletions indicate that constructions can be divided up into their constituent parts, whereas insertions demonstrate that dividing followed by chunking may apply to construction production. Finally, substitutions – which constitute the vast majority of the manipulations seen here – highlight the sophistication of the representations of the constructions that learners acquire. For substitutions, the construction must be abstracted in order to replace a word with an abstract category, into which words of the same category can be inserted. Cartwright and Brent (1997) provided a computational model for how abstractions over these constructions could result from developing an efficient representation that generalises over multiple, similar constructions.

General statistical principles of language acquisition: Grouping and dividing

Data from empirical and observational studies of language acquisition suggest that a range of possible mechanisms may work to generate language structure at different levels, from phonology to syntax, operating on the language environment using statistical computations of varying complexity. These operations can be defined by two broad classifications of mechanistic processing. The first is *grouping*; where operations gather information on which aspects of the language should be chunked together, to construct larger structures from smaller constituents. The second is *dividing*; which operates by determining which larger structures should be divided into their constituent parts. These alternatives may apply differently to language tasks at various levels or may be useful for resolving different aspects of the same language learning task.

For instance, for speech segmentation, computational models typically operate by chunking: determining which phonemes, or phonological features, cohere to form word candidates (e.g., Perruchet & Vinter, 1998). Alternative models have successfully segmented speech corpora into words through dividing; beginning with a larger structure – for instance, a whole utterance – and determining how to divide this up into candidate words (e.g., Monaghan & Christiansen, 2010), or by proposing word boundaries at points where transitional probabilities are low (e.g., Elman, 1993; French et al., 2011). Deciding between these approaches rests on the effectiveness by which these different models account for detailed behavioural data, such as the order as well as the end-point of acquisition, with validity across different languages (Saksida et al., 2017).

For learning word–referent mappings, possible links between words and their meanings may form either through gradually acquiring information about which words and meanings co-occur across multiple learning situations, i.e., by learning to group cross-modal information together. Alternatively, mappings may be learned by determining which links *do not* apply – by applying, for instance, mutual exclusivity, i.e., by learning to divide multimodal information. Thus, a combination of grouping and dividing may reflect word learning behaviour (McMurray, Horst, & Samuelson, 2012). Similarly, for acquisition of morphology, both grouping and dividing appear to be important, with the former being key for production processes, and the latter being more important for storage of the language during acquisition, as implemented in MOSAIC (Freudenthal et al., 2006).

For formation of grammatical categories, statistical mechanisms must identify frequently occurring words in speech, and group words according to the way in which they co-occur with those frequent words, as well as the structural contexts in which they do so. This grouping can be derived from distributional frames in speech (Mintz, 2003; St Clair et al 2010). Finally, for syntax, both grouping and dividing mechanisms appear to operate over utterances to enable language representations to be acquired and used productively (Lieven et al., 2003).

The role of the broader environment on learning

Studies examining the role of statistical learning in language acquisition have primarily focussed on assessing whether and how learners can build linguistic representations from the internal language structure for varying levels of linguistic abstraction. This conventional view is consistent with assumptions about the autonomy of syntax (Jackendoff, 2002), which hold that linguistic representations are modular, and not influenced by non-linguistic processing mechanisms, which apply more generally to non-linguistic stimuli. However, language learners are exquisitely sensitive to multiple cues in their broader language environment, which have been shown to overlap not only with each other, but also with the statistics contained within speech input (Monaghan, Brand, Frost, & Taylor, 2017; Yurovsky et al., 2013). It is therefore unlikely that learners are drawing on distributional information alone during language acquisition, but rather utilise the constellation of cues existing within both the language they hear, and the world around them.

For speech segmentation, there are myriad cues that can help learners to locate word boundaries, including allophonic variation (Salverda et al., 2003), phonotactic constraints (Hockema, 2006), and prosodic information such as stress (Cutler & Norris, 1988) and rhythm (see e.g., Mattys et al., 2005). Corpus studies have highlighted the prevalence of such cues in natural language; for instance, studies of English have shown that stress tends to occur on the first syllable of words (Jusczyk et al., 1993), signalling word onset. Similarly, there is a cross-linguistic tendency for final syllables of words to be lengthened (White & Turk, 2010), indicating word offset in a variety of languages. In addition, phonotactic regularities constrain word boundaries across the world's languages- although the actual constraints critically differ cross-linguistically (for instance, in English "gk" only occurs at word boundaries, as the end of one word and the beginning of another, and words can end but not begin with /ŋ/ (Hockema, 2006), yet this is permissible in other languages (e.g., Gaelic, Khasa).

Whereas statistical information can drive speech segmentation in most laboratory-based studies, there is evidence to suggest that learning is boosted when distributional cues (i.e., associations or transitional probabilities between phonemes or syllables) are supplemented with the additional cues described above. This alliance between transitional probabilities and various other cues may be particularly important given the reported difficulty learners face when (statistically) segmenting words of varying lengths from speech (e.g., Johnson & Tyler, 2010) – as is the requirement in natural language acquisition. Frost and colleagues demonstrated that while adults could segment trisyllabic words using transitional probabilities alone, performance was enhanced when an additional syllable-lengthening cue was used to demarcate the end of a word (Frost et al., 2017, see also Saffran et al., 1996b). Similarly, the interaction between stress cues and statistical information has also been found to influence speech segmentation (Cunillera et al., 2006). In fact, when distributional cues and stress cues conflict, infants from 9 months old have been shown to attend preferentially to stress cues over language statistics (Johnson & Juscyzk, 2001, see also Johnson & Seidl, 2009), with this preference possibly developing over infancy (Thiessen & Saffran, 2003). Saksida et al. (2017) further demonstrated that the rhythmic and distributional properties of speech may critically interact during speech segmentation: Combining corpus analysis with computational modelling, they found systematic differences in how well speech in different languages could be segmented using different statistical learning strategies, with these differences being driven by the rhythmic properties of the languages. Together, these results highlight the fact that computations based on co-occurrence statistics alone are likely to explain only a part of the bigger picture for language acquisition.

For acquisition of word–meaning mappings, a range of cues have been shown to help learning in addition to co-occurrence statistics, including cues that exist both internal (e.g., prosody) and external (e.g., gesture) to the speech signal. Of particular benefit to word learning is the prosodic information contained in speech. The prosodic landscape of the speech children hear is rich; their linguistic input comprises a large quantity of infant-directed speech (e.g., Fernald, 1985; Fernald & Kuhl, 1987), which contains exaggerations of prosodic cues, including more varied pitch, longer durations of vowels sounds in words, shorter utterances, longer pauses, and a lower tempo relative to speech directed toward adults. These cues have been shown to help children learn new words presented within sentences (Ma et al., 2011) as well as in isolation (Graf Estes & Hurley, 2013), possibly by increasing infants' attention during learning. Shukla, White, and Aslin (2011) examined infants' ability to learn from prosodic and distributional cues in combination, and showed that 6-month-olds could segment statistically defined words from utterances and map them to visual referents, but only when a pitch cue aligned with word offset, rather than straddling the word boundary – indicating that prosodic cues may critically interact with distributional cues (i.e., co-occurrence between words and objects in the environment) during word learning.

Infants' word learning has also been found to benefit from gesture (e.g., Baldwin, 1991; Houston-Price et al., 2006), with infants' sensitivity to gestural cues predicting their later vocabulary development (Brooks & Meltzoff, 2008). Eye gaze (Moore et al., 1999) and deictic gestures (Meyer & Baldwin, 2013) have been suggested to assist word learning by directing children's attention toward the referent, and disambiguating between the many possible referents in an array – substantially simplifying the *Gavagai* problem (Quine, 1960) by narrowing the search space. Gesture has also been shown to help learners discern whether a word describes a whole object, or just a part of that object – with whole-object naming more likely to be accompanied by movement by the speaker of the whole (rather than a part) of an object (Gogate et al., 2013). Intriguingly, a recent study demonstrated that adults' word learning was best when learners received a combination of distributional, gestural and prosodic cues that occurred often with referents, but not always – indicating that learning benefitted from not only the combined presence of these cues, but also their imperfect reliability (Monaghan et al., 2017).

For learning dependency-related structures, there is converging evidence to suggest that infant and adult learners can compute over adjacent and non-adjacent items to discover these in speech (Frost & Monaghan, 2016; Marchetto & Bonatti, 2013, 2015; Peña et al., 2002). While perhaps not essential for learning, multiple additional cues may significantly benefit the discovery of non-adjacent structures, by helping to highlight dependencies in the speech stream. For instance, de Diego Balauger and colleagues demonstrated that prosodic cues, specifically pauses between words, can aid learning – indexed both behaviourally (through higher performance on a forced-choice recognition test) and in adults' ERP responses (de Diego-Balauger et al., 2015, see also de Diego Balauger et al., 2007). In addition, Newport and Aslin (2004) suggested that nonadjacent dependency learning may benefit from additional phonological cues, such as phonological similarity between dependent items (see also Frost et al., 2019). Similarly, increasing the salience of dependencies through positioning dependency-carrying syllables toward the edges of sequences (Endress et al., 2005) has also been found to have similar advantages on learning. Together, these cues have been suggested to help learners to select the relevant information that needs to be gathered in order to learn structural regularities involving nonadjacent elements (Rodruiguez-Fornells et al., 2009). Since such learning is possible in the absence of these cues (Frost et al., 2019; Frost & Monaghan, 2016), we suggest they supplement, rather than replace, the distributional statistical mechanism in order to assist learning.

For learning grammatical categories, prosodic (Conwell, 2017) and phonological (Kelly, 1992) grouping cues have been shown to have a profound impact on learning, particularly in addition to information about distributional statistics of categories of words (Monaghan et al., 2005; Monaghan et al., 2007). For instance, Conwell (2017) showed that prosody could be key to processing noun/verb homophones in child-directed speech, with homophones differing in terms of durational cues, pitch, and vowel quality according to the location in which these words were used within utterances. Similarly, in English, disyllabic nouns tend to carry a trochaic stress pattern, while the stress pattern observed in verbs tends to be iambic (Kelly, 1992). Further, English nouns often occur in phrase- or clause-final positions within utterances, and words in these locations are typically longer in duration, providing an additional informative cue for grammatical categorisation.

General cognitive principles of auditory processing have also been shown to support learning for syntactic structure. Nespor and Vogel (1986) demonstrated that during speech production, syntactic structure and prosody tend to align, such that variation in intensity, duration, and pitch can reliably reflect the hierarchical structure of syntax. Prosody has been shown to support syntactic segmentation in infant learners and inform infants' conceptualizations of syntactic constituency (Hawthorne & Gerken, 2014). In adults, prosodic cues (pauses, and intonational pitch contour) have been found to boost learnability of embedded syntactic structures (i.e., centre-embeddings) in artificial grammars, with distributional information alone proving insufficient for learning (Mueller et al., 2010). Trotter et al. (2019) examined the extent to which prosody may inform learners' processing of long-distance dependencies in complex syntactic structures. They analysed adults' utterances on a prior task eliciting relative clause productions (Montag & MacDonald, 2014), and demonstrated that adults' use of prosody varied systematically depending on whether the relative clauses they produced were active versus passive (Deutsch, 2013; Fery & Schubo, 2010) – providing low-level auditory grouping cues that may support syntactic processing.

Adults have also been shown to draw on prosodic cues for syntactic disambiguation, with cue-use varying cross-linguistically. For instance, bilingual speakers of English and German (both English-German, and German-English) were able to use pitch rise and pitch accent to parse ambiguous prepositional phrase-attachment structures in German, yet used only pitch-accent for processing comparable sentences in English – in line with the properties of these languages (O'Brien et al., 2014).

The broad array of environmental information can be accommodated easily into a statistical learning framework, wherein the learner manipulates any information that is useful to the task at hand, regardless of its modality of presentation. This view contends that language learning mechanisms are not autonomous from other perceptual and cognitive mechanisms (e.g., Newmeyer, 2017). Multiple cues have been found to assist with language acquisition for a range of tasks, which vary in their complexity. Crucially, infants may draw on these language cues in combination, however the use of particular cues, and the way these interact, may change over time as children develop (see e.g., Hollich et al., 2000). Empirical, observational, and computational studies have demonstrated that these multiple cues may work together differently across languages (Saksida et al., 2017), with learners developing strategies for cue-use that align with the topographical distribution of those cues in the input; giving greater weighting to cues that are more available, and more reliable within that language (e.g., MacWhinney et al., 1984). Thus, while the importance of distributional statistics (for phonetic features, phonemes, syllables, and words) for learning is indisputable, this information is not likely to be processed in isolation during language acquisition. In fact, use of cues is likely to be complex, dynamic, and varying for speakers across development, and across the world's languages.

A note on cue variability

A fundamental aspect of distributional language acquisition is the role of variability in learning. The language environment, and the multimodal landscape in which language is situated, is replete with noise. For instance, myriad cues are present in speech that may help with speech segmentation, but these cues seldom occur with perfect reliability; learners could look to pauses, but these only sometimes intervene between words, and often only at phrasal boundaries. Equally, while prosodic cues are pervasive, they are not absolute; English verbs often carry stress on the second syllable, countering the trend for word-initial stress, whereas some affixes attract stress at word-medial and word-final locations. Similarly, while gestural cues can guide infants toward a referent, they can sometimes be misleading (for instance, if a caregiver is playing with the child while also talking about something else). Thus, natural language distribution is rather more chaotic than is typically accounted for in empirical studies, and it is conceivable that this noise may have a substantial impact on language acquisition.

Humans have been shown to be capable of processing language robustly against sometimes substantial environmental noise. In fact, an emerging theme in the literature is that variability within language is actually advantageous for learning. Variability in phoneme productions has been shown to act as an additional source of data for predicting likely up-coming words (McClelland & Elman, 1986) as well as word-boundaries (Salverda et al., 2003). In machine learning studies, a pristine training environment, without any noise or variation in the environmental input, can result in brittle learning that is unable to effectively generalise to other situations (Ay, Flack, & Krakauer, 2007), with stable learning resulting from variability either within the environment (Monaghan, 2017) or within the system itself (Whitacre, 2010). A recent computational model of word learning suggests this noise may serve to help, rather than hinder, learning, by guiding the creation of a robust, canalised language system that is vitally resistant to noise in the learning environment (Monaghan, 2017). The value of variability – and the benefit for computational models of learning from multiple, probabilistic cues – is also consistent with avoiding blocking effects in associative learning studies (Nixon, submitted).

Empirical studies examining the impact of environmental noise on learning have demonstrated benefits of variability in the speech signal for a number of linguistic tasks. For instance, a recent study found that statistical segmentation of target words from speech was best when targets occurred alongside high frequency words often, but not always, in a continuous stream (Frost et al., in prep). Similarly, the Zipfian distribution (Zipf, 1935) of words within language has been found to significantly benefit speech segmentation (Kurumada et al., 2013) – indicating that the varied use of particular words may be advantageous for learning. Equally, variability within the distributional statistics of language has been found to have advantages for word learning (Hendrickson & Perfors, 2019; Monaghan et al., 2017), semantic category learning (Lany, 2014), and syntactic structure acquisition (Gomez, 2002). Variation in the availability and reliability of distributional cues may advantage learning by encouraging learners to seek multiple potential information sources to which they can look for guidance, lessening the importance of a particular individual cue, and increasing the resilience of the language system to noise.

Yurovsky et al. (2013) demonstrated that infant learners are particularly adept at drawing on probabilistic cues occurring within the environment. Further, they demonstrated that infants' expert capacity for cue use is able to overcome problems of cue-dilution demonstrated for adult learners, whereby predictions reduce in accuracy when drawing on cues of different strengths, compared to just a single cue. The Intersensory Redundancy Hypothesis (Bahrick et al., 2004) posits that cue co-occurrence increases the saliency of individual cues by guiding learners to other informative cues and reinforcing their perceived usefulness. While this possibility is intriguing, this account fails to explain the way cues have been shown to assist with other tasks in language acquisition, particularly concerning the way that cues may interact during learning, or the way that cue use may change over language development. A more comprehensive explanation of how learning may take place may instead lie in accounts which posit the mechanisms underlying language acquisition form a dynamic system, wherein attention integrates with statistical processing of co-occurrences between potential referents in the visual and sensory environment and words in the language environment.

Conclusions

Children have an extraordinary capacity to draw on the distributional properties of speech and have been shown to apply this ability to learn about a broad range of linguistic features of varying complexity. Studies of statistical learning have helped us determine the nature of the processes underlying the necessary computations for different tasks in language acquisition and have shed light on the way that children use these processes to operate over speech input on their journey to linguistic proficiency. Results from studies exploring the way that learners can draw on a range of cues (within speech, and in the infants' wider environment) alongside distributional information indicate that learners use many cues in combination during language acquisition, with the interaction between particular cues complex and dynamic; determined by the properties of the language, and the environment in which it is being learnt, with variation of cue availability and reliability influencing this interaction further still.

Funding

This work was supported by the International Centre for Language and Communicative Development (LuCiD). The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

References

Aristotle (350BCE, 1932). Posterior analytics. Translated by G. R. G. Mure, *Works of Aristotle, Volume 2*. Oxford: Oxford University Press.

Ay, N., Flack, J., & Krakauer, D. (2007). Robustness and complexity co-constructed in multimodal signalling networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1479), 441–447. https://doi.org/10.1098/rstb.2006.1971

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13, 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 124–129). Austin, TX: Cognitive Science Society.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, 35(1), 207–220. https://doi.org/10.1017/S030500090700829X

Cameron-Faulkner, T., Lieven, E. V., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, 63(2), 121–170. https://doi.org/10.1016/S0010-0277(96)00793-7

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36, 1–22. https://doi.org/10.1162/0024389052993655

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B*, 372 (1711), 1–10. https://doi.org/10.1098/rstb.2016.0055

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, 63(3), 522–543. https://doi.org/10.2307/415004

Conwell, E. (2017). Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Chid Language*, 44(3), 734–751. https://doi.org/10.1017/S030500091600009X

Culter, A., & Norris, D.. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, 14, 113–121.

Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodruiguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, 1123(1), 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

De Diego-Balaguer, R., Rodriguez-Fornells, A. & Bachoud-Lévi, A. C. (2015). Prosodic cues enhance rule learning by changing speech segmentation mechanisms. *Frontiers in Psychology*, 6, 1478. https://doi.org/10.3389/fpsyg.2015.01478

De Diego-Balauger, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Levi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One*, 2, 01175.

Deutsch, D. (2013). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 184–238). San Diego, CA: Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00006-7

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, 173, 43–59. https://doi.org/10.1016/j.cognition.2017.11.008

Elman, J. L. (1993). Learning and developmentin neural networks: The importance of starting small. *Cognition*, 48, 71–99.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Psychology: General*, 134(3), 406–419.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, 8(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, 10(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fery, C. & Schubö, F. (2010). Hierarchical prosodic structures in the intonation of center embedded relative clauses. *The Linguistic Review*, 27(3), 293–317. https://doi.org/10.1515/tlir.2010.011

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., & Gobet, F. (2006). Modeling the development of children's use of optional infinitives in Dutch and English using MOSAIC. *Cognitive Science*, 30, 277–310. https://doi.org/10.1207/s15516709cog0000_47

Fries, C. C. (1952). *The structure of English*. London: Longmans.

Frost, R. L. A., Isbilen, E. S., Christiansen, M. H. & Monaghan, P. (2019). Testing the limits of non-adjacent dependency learning: Statistical segmentation and generalization across domains. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. Montreal, QB: Cognitive Science Society.

Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., C., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, 120: 101291.

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Frost, R. L. A., Monaghan, P., & Tatsumi, T. (2017). Domain-general mechanisms for speech segmentation: The role of duration information in language learning. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 466–476.
Gogate, L. J., Maganti, M., & Laing, K. (2013). Maternal naming of object wholes versus parts for preverbal infants: A fine-grained analysis of scaffolding at 6 to 8 months. *Infant Behavior & Development*, 36(3), 470–479. https://doi.org/10.1016/j.infbeh.2013.03.012
Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436. https://doi.org/10.1111/1467-9280.00476
Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5), 797–824. https://doi.org/10.1111/infa.12006
Harris, Z. S. (1954). Distributional structure. *Word*, 10, 140–162. https://doi.org/10.1080/00437956.1954.11659520
Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31, 190–222. https://doi.org/10.2307/411036
Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, 133, 420 – 428. https://doi.org/10.1016/j.cognition.2014.07.013
Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition*, 189, 11–22. https://doi.org/10.1016/j.cognition.2019.03.005
Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development*, 2, 119–146. https://doi.org/10.1207/s15473341lld0202_3
Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, 65(3, Serial No. 262).
Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, 95, 27–55. https://doi.org/10.1016/j.jecp.2006.03.006
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–567. https://doi.org/10.1006/jmla.2000.2755
Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, 12(1), 131–141. https://doi.org/10.1111/j.1467-7687.2008.00740.x
Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x
Jusczyk, P. W., Cutler, A., Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64(3), 675–687. https://doi.org/10.2307/1131210
Kamper, H., Jansen, A., & Goldwater, S. (2016). Unsupervised word segmentation and lexicon discovery using acoustic word embeddings. In *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24, 669–679. https://doi.org/10.1109/TASLP.2016.2517567
Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349–364. https://doi.org/10.1037/0033-295X.99.2.349

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

Lany, J. (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development*, 85(4), 1727–1739.

Lieven, E. V., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30(2), 333–370. https://doi.org/10.1017/S0305000903005592

Lieven, E. V., & Brandt, S. (2011). The constructivist approach. *Infancia y Aprendizaje*, 34(3), 281–296. https://doi.org/10.1174/021037011797238586

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

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*,7(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

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

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behaviour*, 23, 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology 67*(3), 130–150. https://doi.org/10.1016/j.cogpsych.2013.08.001

Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, 42(4), 873–902. https://doi.org/10.1017/S0305000914000452

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877. https://doi.org/10.1037/a0029872

Meyer, M., & Baldwin, D. A.. (2013). Pointing as a socio-pragmatic cue to particular vs. generic reference. *Language Learning and Development*, 9(3), 245–265. https://doi.org/10.1080/15475441.2013.753802

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91–117. https://doi.org/10.1016/S0010-0277(03)00140-9

Modrak, D. K. W. (2001). *Aristotle's theory of language and meaning*. Cambridge: Cambridge University Press.

Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science*, 9, 21–34.

Monaghan, P., Brand, J., Frost, R. L. A., & Taylor, G. (2017). Multiple variable cues in the environment promote accurate and robust word learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 817–822). Austin, TX: Cognitive Science Society.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition*, 96, 143–182. https://doi.org/10.1016/j.cognition.2004.09.001

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, 37, 545–564. https://doi.org/10.1017/S0305000909990511

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional Coherence Hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55, 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Monaghan, P. & Mattock, K. (2012). Integrating constraints for learning word referent mappings. *Cognition*, 123, 133–143. https://doi.org/10.1016/j.cognition.2011.12.010

Monaghan, P., Mattock, K., Davies, R., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, 39, 1099–1112. https://doi.org/10.1111/cogs.12186

Monaghan, P., Kalashnikova, M., & Mattock, K. (2017). Intrinsic and extrinsic cues to word learning. In G. Westermann & N. Mani (Eds.), *Early word learning*. Hove: Psychology Press. https://doi.org/10.4324/9781315730974-3

Moore, C., Angelopolous, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, 35(1), 60–68. https://doi.org/10.1037/0012-1649.35.1.60

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science*, 34(2), 338–349. https://doi.org/10.1111/j.1551-6709.2009.01093.x

Nespor, M. & Vogel, I. (1986). Prosodic Phonology. Dordrecht: Foris Publications

Newmeyer, F. J. (2017). Form and function in the evolution of grammar. *Cognitive Science*, 41, 259–276. https://doi.org/10.1111/cogs.12333

Newport, E. L., & Aslin, R. (2004). Learning at a distance: Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Nixon, J. S. (submitted). Of mice and men: Is speech sound acquisition statistical or error- driven?

O'Brien, M. G., Jackson, C. N., Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Applied Psycholinguistics*, 35(1), 27–70. https://doi.org/10.1017/S0142716412000252

Pelucchi, B., Hay, J. F., Saffran, J. R. (2009). 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., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning non- adjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology*, 133(4), 573–583). https://doi.org/10.1037/0096-3445.133.4.573

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607. https://doi.org/10.1126/science.1072901

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pullum, G. K., & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 19, 9–50.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: The MIT Press
Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic structures. *Cognitive Science*, 22, 425–469. https://doi.org/10.1207/s15516709cog2204_2
Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., & De Diego-Balauger, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 364(1536), 3711–3734. https://doi.org/10.1098/rstb.2009.0130
Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928. https://doi.org/10.1126/science.274.5294.1926
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. https://doi.org/10.1006/jmla.1996.0032
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8 (2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x
Saksida, A., Langus, A., & Nespor, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science*, 20(3), e12390. https://doi.org/10.1111/desc.12390
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2
Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, 122, 163–180. https://doi.org/10.1016/j.cognition.2011.10.010
Shukla, M., White, K. S., & Aslin, R. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *PNAS*, 108(15), 6038–6043. https://doi.org/10.1073/pnas.1017617108
Smith, L., & Yu, C. (2008). Infants rapidly learn word–referent mappings via cross-situational statistics. *Cognition*, 106, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010
Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004
St Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, 116, 341–360. https://doi.org/10.1016/j.cognition.2010.05.012
Stumper, B., Bannard, C., Lieven, E. V., & Tomasello, M. (2011). "Frequent frames" in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science*, 35(6), 1190–1205. https://doi.org/10.1111/j.1551-6709.2011.01187.x
Swingley, D. (2005). Statistical clustering and the contents of infant vocabulary. *Cognitive Psychology*, 50(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

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, 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Trotter, A. S., Frost, R. L. A., & Monaghan, P. (2019). Chained melody: Low-level acoustic cues as a guide to phrase structure in comprehension (Unpublished doctoral dissertation).

Whitacre, J. (2010). Degeneracy: A link between evolvability, robustness and complexity in biological systems. *Theoretical Biology and Medical Modelling*, 7, 6. https://doi.org/10.1186/1742-4682-7-6

White, L., & Turk, A. E. (2010). English words on a Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38(3), 459–471. https://doi.org/10.1016/j.wocn.2010.05.002

Yu, C., & Smith, L. B. (2012). Modeling cross situational word referent learning: Prior questions. *Psychological Review*, 119(1), 21–39. https://doi.org/10.1037/a0026182

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, 16, 959–966.

Yurovsky, D., Boyer, T. W., Smith, L. B., & Yu, C. (2013). Probabilistic cue combination: Less is more. *Developmental Science*, 16(2), 149–158. https://doi.org/10.1111/desc.12011

Zipf, G. K. (1935). *Psycho-biology of languages*. Cambridge, MA: The MIT Press.