16. Statistical learning as a domain-general mechanism of entrenchment

Ethan Jost & Morten H. Christiansen

Abstract

Statistical learning is a cognitive process that serves as a mechanism of entrenchment across a number of domains, including language. It is a process by which learners implicitly form associations between stimuli by tracking and storing the underlying statistical relationships between such elements. This chapter examines the statistical relationships that learners are sensitive to, along with the nature of these relationships within the various modalities in which statistical learning has been studied. The automaticity and implicit nature of statistical learning is discussed, in addition to the relative contributions of statistical learning to language development, from phonology to grammar. The literature on individual differences in statistical learning also serves to elucidate the relationship between statistical learning and language. Finally, models of language acquisition relying on the mechanism of statistical learning are presented as further evidence for the importance of statistical learning in understanding the cognitive basis of language learning and processing.

1. Introduction

The birth of the statistical learning literature is often traced back to Reber's (1967) seminal study on implicit learning using an artificial grammar learning paradigm. However, to fully understand the relationship between such early implicit learning studies and the current notion of statistical learning, it is important to also consider its conception. The theory of perceptual learning by Gibson and Gibson (1955) paved the way for accounts of learning with a basis in sensory experience. In the Gibsons' theory of perceptual learning, which has close parallels to current ideas about entrenchment (Schmid, 2007), repeated experience with a percept enhances one's ability to discriminate between it and other percepts. This chapter will argue that a communicative system characterized by entrenchment, as posited in this volume, likely relies to a considerable extent on the ability to track, learn, and use underlying associative relationships between linguistic elements and structures in comprehension and production.

When considering the origin of statistical learning as a theoretical construct, it is also important to consider the early work of Miller and Selfridge (1950) who thought that a reliance on transitional probabilities may be similar to the way in which grammar is learned. Other research informed by both Miller's work and the theory of perceptual learning espoused by the Gibsons' demonstrated that frequent co-occurrence due to underlying structure improved participants' recall of letter sequences (Miller, 1958), and that learning the positional relationships between linguistic units (i.e., morphemes) occurs as an experiential process of familiarization with the temporal positions in which such units are frequently encountered (Braine, 1963). This laid the foundation for future research investigating the close relationship between frequent co-occurrence and the strength and automaticity of recall at various levels of linguistic analysis.

From the beginning, research on implicit learning related to language was focused on the way(s) in which units of linguistic information are formed. Some of the early explanations for the ways in which this learning happened relied upon experience-based accounts, as described above. However, experience-independent theories of language acquisition quickly became the dominant perspective primarily due to the widespread acceptance of the "poverty of the stimulus" argument (Chomsky, 1965; Crain, 1991). Saffran, Aslin and Newport's (1996a) research gave the psychology of language an experience-dependent statistical learning mechanism by which at least one aspect of linguistic knowledge (words) could be learned, and demonstrated that this could be accomplished fairly rapidly even at an early stage in development; statistical learning can thus be thought of as the acquisition of distributional information from perceptual input.

While the exact nature of the distributional information learners are thought to be sensitive to varies from study to study, this chapter aims to bring together research from multiple perspectives, in order to give a thorough overview of the field. The kinds of statistics that learners are using in each task and study will be highlighted and contrasted, particularly when such differences are important from a theoretical standpoint. With the uncovering of this learning mechanism and the increased weight given to connectionist ideas about how the items and structure of language can emerge from the input (Elman, 1990), experience-dependent accounts of language learning and processing have again become central to the psychology of language. Building these ideas we define statistical learning[1] for the purpose of this chapter as the process by which learners uncover the structure of the input from its distributional properties (Frost, Armstrong, Siegelman, & Christiansen, 2015).

1.1. Implicit learning meets statistical learning

Since the resurgence of experience-dependent accounts of language in the 1990s, attempts have been made to synthesize the original implicit learning literature with the newer research on statistical learning (e.g., Conway & Christiansen, 2006; Perruchet & Pacton, 2006). Researchers have begun to question the "implicitness" of statistical learning, and the related artificial grammar paradigms that are common within the implicit learning literature. This is particularly relevant to discussions of entrenchment processes, as automaticity – or unconscious activation – is usually considered a feature of entrenchment (Schmid, 2007; see Moors and Hartsuiker, this volume, for more details); the naming of an entrenched visual stimulus (i.e., an apple) does not require conscious processing in healthy adults. However, considering the manner in which most statistical learning paradigms are designed, with explicit familiarity judgments used at test, the relative amount of conscious processing that learners rely upon has been debated.

[1] Note that the term "statistical learning" means something quite different in psychology than it does in the field of mathematics and machine learning (Vapnik, 1999). Also, there a number of other learning theories within psychology that are neither at odds with statistical learning, nor do they necessarily fall under the same umbrella, such as discriminative learning (Baayen, 2010). Such ideas about contextual learning can rather be thought of as parallel processes that also help to explain the way that learners gain knowledge from input, in conjunction with cognitive mechanisms like statistical learning.

Within most statistical learning studies, self-report data and the mere fact that the instructions are incidental are used as evidence for implicit processing. Recent work has put this to the test, with evidence both for (Kim, Seitz, Feenstra, & Shams, 2009) and against (Bertels, Franco, & Destrebecqz, 2012) implicit interpretations of statistical learning. Further research has shown that access to the statistical relationships within two artificial languages can be consciously controlled, demonstrating that at least some aspects of the learned relationships is available for explicit processing (Franco, Cleeremans, & Destrebecqz, 2011). Early artificial grammar learning research pointed towards diminished performance when participants were given explicit instructions (Reber, 1976), although newer research suggests that the duration of stimulus presentation may modulate this relationship, with longer presentations leading to an improvement in learning when instructions are explicit, at least in the visual domain (Arciuli, Torkildsen, Stevens, & Simpson, 2014). There appears to be a strong argument for the implicit and incidental nature of statistical learning, but some room for explicit processing should be built into accounts of statistical learning. Some of the issues in understanding the implicit nature of statistical learning are due to the lack of coherence between the implicit and statistical learning literatures, but may be resolved in time as the two become more closely integrated.

Perruchet and Pacton (2006) have claimed that while the two literatures have grown increasingly similar in terms of methodology, implicit learning relies more on the process of chunking as an explanation of learning (see Gobet, this volume), while the statistical learning literature is primarily interested in exploring the role of distributional information. However, these computations do not need to be interpreted as dichotomous; depending on the properties of the input they could both occur in what we think of as statistical learning (Franco & Destrebecqz, 2012).

Tracking conditional probabilities may lead to the formation of chunks at later stages of learning, which then become elements themselves between which conditional probabilities may be tracked. In fact, recent models of language acquisition have demonstrated the feasibility of such a process (Monaghan & Christiansen, 2010; McCauley & Christiansen, 2014). Thinking of chunks as the outcome of statistical learning provides a direct connection with entrenchment: Throughout learning, frequently co-occurring elements and structures become more deeply entrenched, strengthening such representations.

1.2. Statistical learning as a mechanism of entrenchment

This perspective fits in nicely with the notion of entrenchment in language, and promotes the idea of statistical learning as a mechanism of entrenchment. Entrenchment itself is often thought of as a process, but it can also be viewed as an effect. In this way, statistical learning can itself be thought of as part of the process by which entrenchment can occur. The well-established effect of frequency on processing linguistic elements and structures (Oldfield & Wingfield, 1965) can be viewed as a measure of entrenchment in language, although new, more sensitive measures such as meaning-dependent phonetic duration and reading time effects may lead to a more nuanced view of the entrenchment process (Jolsvai, McCauley, & Christiansen, 2013). Therefore, a continuously updated relationship due to the tracking of distributional information and associated formation of meaningful units can lead to varying degrees of entrenchment for any particular element. This interpretation of entrenchment would relate to the learning of a word from a continuous stream of speech (Saffran et al., 1996a) and to the formation of chunks including frequently co-occurring non-adjacent morphemes (Gómez, 2002), along with other linguistic structures.

However, statistical learning is not a mechanism of entrenchment solely in the linguistic domain. In the auditory domain it may also pertain to the learning of tone sequences (Saffran, Johnson, Aslin, & Newport, 1999), while in the visual domain it may relate to the extraction of probabilistic information and structure from visual scenes (Fiser & Aslin, 2002). This points to another key aspect of statistical learning, specifically, its domain-general nature. Statistical learning can take place between stimuli within various sensory modalities (Conway & Christiansen, 2005), and statistical relationships between actions, labels, and referents can be tracked across situations (Yu & Smith, 2007). Along with other domain-general cognitive processes including attention, memory, communicative inference, and general world knowledge, we can understand language as being built upon a foundation that is not specific to language (for a detailed overview of this perspective, see Christiansen & Chater, 2008). Understanding statistical learning as domain-general is also important for considering the ways in which language and statistical learning interact with other aspects of cognition.

2. Statistical learning in multiple domains

The domain-generality of statistical learning has been extensively studied since the advent of the modern statistical learning literature. This aspect of the statistical learning mechanism is important for a number of reasons. To begin with, it tied into assumptions about implicit learning, proposed by Reber (1993), who hypothesized that implicit learning was a phylogenetically ancient and conserved cognitive ability. Given that other species possess complex communication but not language, this meant that artificial grammars with non-linguistic elements ought to be learnable by humans, and likely some other extant species. However, strong theories of cognitive modularity argue that the cognitive architecture is built out of domain-specific modules (Fodor, 1983). Thus, experimental findings in which similar cognitive processes were utilized by different hypothesized modules (e.g., between vision and language) provided counter-evidence to such claims. Due primarily to these theoretical motivations a number of researchers have attempted to elucidate the extent of the generality of this mechanism.

2.1. Statistical learning at multiple levels of linguistic processing

The first studies of statistical learning focused on the learnability of word-like units from a continuous stream of syllables based solely on the different transitional probabilities within vs. between “words” in infants (Saffran et al., 1996a) and adults (Saffran, Newport, & Aslin, 1996b). In adults it was found that additional prosodic cues at word boundaries facilitated learning.

Within the statistical learning literature, this type of relationship between syllables would come to be defined as an adjacent dependency (e.g., */tu-pi-ro/da-pu-ki/)*. It was suggested to be analogous to the type of statistical relationship formed between syllables within words vs. between words in terms of lexical processing; the syllable transitions that are found within words (*pi-ro*) have higher transitional probabilities than the syllable transitions that exist between words (*ro-da*). The conditional probabilities that are tracked between adjacent items in a sequence lead to the learning of these frequently co-occurring items.

Another type of dependency that has become part of statistical learning parlance is the non-adjacency (e.g., *a/X/d* where ‘a’ predicts ‘d’ with various random intervening elements instantiating ‘X’) (Gomez, 2002). The non-adjacent dependency in statistical learning paradigms was argued to be similar to the type of relationship found between auxiliaries and inflectional morphemes (e.g., *was running*; *had beaten*) and number agreement across multiple words (e.g., *the dogs out in the yard are howling*).

These studies, when combined with the research on adjacent dependencies, point to powerful learning mechanisms that may underlie entrenchment across a variety of linguistic domains. That is, learners seem to be sensitive to continuously updated statistical/probabilistic relationships not just between items that are temporally adjacent, but also across intervening items, so long as the intervening items are sufficiently variable. Additional evidence from the ERP literature has demonstrated that the brain processes syllable-to-syllable transitions in Saffran-style statistical learning paradigms differently within vs. between words, as greater N100 amplitudes were found at between-word syllable boundaries than at within-word syllable boundaries (Sanders, Newport, & Neville, 2002). The N100 is often thought to reflect early bottom-up sensory processing (van den Brink, Brown, & Hagoort, 2001).

2.2. Statistical learning in different domains

It is important to note that the basic units of learning (syllables) in these statistical learning paradigms are the same as what are thought of as one of the most basic units of language, thus these non-word stimuli are typically described as linguistic in nature (Newport & Aslin, 2004). However, the stimuli used in statistical learning paradigms are not limited to language-like items. Statistical learning has been studied in a number of other domains, including audition, vision, and touch.

2.2.1. Audition

If statistical learning was domain-specific and only related to the way in which language is learned and processed, then statistical relationships between non-linguistic elements should not be learnable. This appears not to be the case, as the ability to learn from the transitional probabilities in sequences of auditory tones has been well described in the literature. Saffran and colleagues (1999) first reported the sensitivity of adults and infants to the underlying statistical relationships between tones, using the same type of dependency previously investigated using syllables (Saffran et al., 1996a, 1996b). The ability of participants to track adjacent dependencies between tones that are inherently non-linguistic indicates that statistical learning is likely a domain-general mechanism.

Other kinds of acoustic information have also been used in statistical learning studies, with varying results depending on the properties of the acoustic stimuli (Creel, Newport, & Aslin, 2004). Interestingly, certain aspects of the stimulus (e.g., pitch register and timbre) led to different patterns of sensitivity in learning non-adjacency vs. adjacency structure in the stimulus stream, suggesting that Gestalt-like properties of the stimulus may shape learning in different ways. Other reports of statistical learning have relied on artificial grammars using musical stimuli, further demonstrating the domain-general nature of statistical learning (e.g., Bly, Carrion, &

Rasch, 2009). This domain-generality indicates that language is subserved by neural mechanisms that are used for processing a variety of input, and/or that the same general computational principle operates across perceptual and cognitive domains.

2.2.2. Vision

Auditory input is still somewhat language-like, as that sensory modality is used for listening to speech. Vision is a sensory domain further removed from language processing, and to find that statistical learning of visual sequences is possible would strengthen claims about this mechanism's domain-general nature. Evidence of visual statistical learning began with a study examining infant looking times to statistically determined patterns of shapes, finding differences in looking times between familiar and unfamiliar patterns (Kirkham, Slemmer, & Johnson, 2002; Fiser & Aslin, 2002). The statistical coherence between elements within these visual scenes led to their entrenchment as higher-order representations. The features of visual stimuli often consist of color, shape, and positional information with various types of biases existing between learning these features vs. objects (Turk-Browne, Isola, Scholl, & Treat, 2008), similar to the effect of the stimulus-level differences noted in auditory statistical learning. For example, when two features, such as color and shape, perfectly co-vary within each object in a triplet, participants struggle to identify acceptable triplets when tested on only one of the two features (either color or shape). However, when shape and color are decoupled during training and vary across objects, the underlying pattern for each feature can be learned independently. In terms of development, adults and children seem to show similar underlying neural processes when learning sequential information in the visual domain, with stable P300 responses across age groups to visual stimuli that are highly predictive of a target stimulus (Jost, Conway, Purdy, Walk, & Hendricks, 2015).

2.2.3. Touch and other domains

Touch is another modality in which statistical learning has been studied. Conway and Christiansen (2005) investigated whether or not statistical structure could be learned purely from tactile input. They found that performance with tactile input is similar to performance in the visual modality, though auditory learning was superior to both when the same artificial grammar was used in each modality. Further theories point towards the use of a statistical learning mechanism as a basis for social understanding (Lieberman, 2000; Ruffman, Taumoepeau, & Perkins, 2012) and motor skill learning (Robertson, 2007).

These findings lead to interesting questions about what kinds of constraints are placed on learning due to the nature of stimuli in different sensory modalities. For example, auditory information is usually encountered in rapid succession and is quite transient in nature. Thus, basic sensory processing mechanisms for auditory input are tuned to this bias in presentation. Visual input varies across time as well, but is much more stable and thus statistical learning studies incorporating visual stimuli require longer inter-stimulus intervals to achieve the same levels of learning as in audition (Emberson, Conway, & Christiansen, 2011). One possible explanation for the patterns of similarity and differences in statistical learning across domains is the existence of multiple modality-specific mechanisms, each using the same underlying computational principles, but subject to different modality-specific constraints (Frost et al., 2015).

The evidence of statistical learning across different modalities and domains suggests that entrenchment might not be a language-specific phenomenon. Examples such as the incidental categorization of single tones into triplets due to frequent co-occurrence in a continuous stream (e.g. Saffran et al., 1999) and the extraction of statistical structure from visual scenes (e.g. Fiser & Aslin, 2002, Kirkham et al., 2002) provide compelling arguments for statistical learning as a domain-general process of entrenchment. The construction of holistic units out of basic elements is a hallmark of entrenchment. Building tone triplets out of a sequence of single tones based on co-occurrence may not be perfectly analogous to the process by which linguistic structures are thought to be entrenched, but it does capture the basic properties of a process which, as described above, may operate at various levels of linguistic processing as a foundation for the formation of such associations.

3. Statistical learning in development

This section will focus on developmental changes in statistical learning abilities, and how such changes may affect language development (for an extended review, see Misyak, Goldstein, & Christiansen, 2012). The human infant is born into a world full of input from which it must extract structure (James, 1890; Goldstein et al., 2010). While this may seem to be a difficult task, the infant's environment, experience, and biology constrain the kinds of input to which it is sensitive (Elman et al., 1996). However, actually extracting structure from the input requires some kind of learning mechanism; this is where statistical learning comes into play.

Reber (1993) hypothesized that implicit learning was developmentally invariant due to its basic adaptive value and ancient phylogenetic roots. Therefore, if Reber were correct, robust statistical learning mechanisms should be present from an early age. Indeed, infant studies formed the foundation for modern research on statistical learning, as humans seem to possess powerful statistical learning abilities from infancy (Saffran et al., 1996a; Saffran et al., 1999). By at least eight months, infants can track an aspect of the speech stream which allows them to learn words, and appear to do so in a way similar to adults (Saffran et al., 1996b). In the domain of vision, older children from ages six to twelve have been found to possess neural correlates of learning similar to adults in a simple sequential learning paradigm, giving credence to Reber's claim in a non-linguistic task (Jost et al., 2015). Amso and Davidow (2012) have also provided compelling evidence for developmental invariance in statistical learning of environmental regularity by examining saccadic eye movements and reaction times to probabilistically determined object relationships in infants and adults.

Deeper investigation into the developmental invariance of statistical learning has provided some counter-evidence, forcing a reappraisal of Reber's original position. From birth, infants have the ability to segment continuous speech using statistical information, as evidenced by ERPs (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). However, infants and adults may be sensitive to different components of auditory information than adults, as infants do not track statistical relationships defined by relative pitch whereas adults do (Saffran & Griepentrog, 2001). In another study by Saffran (2001), adults and children both performed above chance on measures of learning following exposure to an artificial grammar containing predictive dependencies, but adults consistently outperformed children. While this may have been due to differences in memory ability, as children consistently performed worse on longer strings while adults did not show the same effect, the influence of other cognitive processes on statistical learning ability along with widely varying amounts of experience with stimuli in various domains (i.e., sensitivity to relative pitch) may contribute to developmental differences in statistical learning abilities.

The first study of non-adjacency learning in adults and infants also found that infants possess adult-like abilities to track such dependencies (Gomez, 2002). However, developmental differences in the ability to learn from non-adjacent dependencies have also been found (Gomez & Maye, 2005). Twelve-month old infants were unable to learn the predictive relationship between non-adjacent elements in a task that infants fifteen-months and older were able to perform. At this point, it seems likely that true developmental invariance is not a characteristic of statistical learning and that studies reporting such findings do not include a sufficient range of ages across development.

This growing literature on statistical learning in development not only demonstrates the existence of statistical learning abilities at early stages of development, but also provides a window into the interaction between experience and cognitive development. It seems clear that infants have access to cognitive mechanisms that contribute to the entrenchment of lexical items as well as certain aspects of linguistic structure. However, sensitivity to certain statistical properties of speech from the very onset of development as opposed to others may bias the types of language learning we see in development. Considering that neonates have the ability to learn syllable chunks by tracking adjacent dependencies, a mechanism for the construction of lexical items seems to exist very early in development (Teinonen et al., 2009).

The idea that humans, and infants in particular, are guided by statistical structure when learning, due to a fundamental attempt to reduce uncertainty (Gibson, 1991; Gomez, 2002), provides an explanation for the way in which language develops. Sensitivity to the adjacent structures in language provides quite a bit of information, and allows for syllables to become associated with one another to form words, and for words to become associated with one another, forming chunks, while sensitivity to the non-adjacent structures in language provides a means by which more complex associations required for learning certain aspects of morphology and syntax, for example constructional schemas, are developed. In this way, statistical learning contributes to entrenchment of both linguistic elements, and linguistic structures.

4. Individual differences in statistical learning and language

Reber (1993) stated that due to the fundamentally ancient nature of implicit learning it was unlikely that there would be profound individual variation in related abilities. While he has since reconsidered his claim (Reber & Allen, 2000), his initial hypothesis has had a great deal of influence on the field of statistical learning. However, recent evidence has pointed towards individual variation in statistical learning abilities, and studies of this evidence have also attempted to elucidate how these individual differences contribute to differences in language abilities (see Frost et al., 2015, for a discussion).

Shafto, Conway, Field, and Houston (2012) have provided developmental evidence for direct links between individual differences in statistical learning and language abilities. Prelinguistic infants aged eight-and-a-half months had their learning abilities evaluated on a visuospatial statistical learning task, and then five months later were assessed for their early language skills using the MacArthur-Bates Communicative Development Inventory. Early statistical learning abilities were found to predict language development, as infants who were able to track the statistical relationships in the visual learning paradigm showed better language outcomes than those who did not. More longitudinal studies investigating the relationship between statistical learning and language would greatly benefit our understanding of their relationship (Arciuli & Torkildsen, 2012).

Other individual differences studies with adult participants have demonstrated co-variation between statistical learning and language abilities. One study found that individuals' performance on a visual statistical learning task was correlated with performance on a task designed to test linguistic knowledge by querying whether or not they were able to decipher a predictable word in degraded auditory conditions (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Individuals' statistical learning scores have also been found to be a better predictor of language comprehension than performance on a verbal working memory task (Misyak & Christiansen, 2012). Another study in which implicit learning was identified as a distinct cognitive ability found it to be associated with verbal analogical reasoning (Kaufman et al., 2010).

The previous discussion of adjacency and non-adjacency learning has painted a picture of two similar computations performed over the same kinds of stimuli but with varying spatio-temporal signatures. It seems plausible that they contribute to the entrenchment of linguistic features at multiple levels of processing, and are recruited preferentially depending on the structure of the statistical relationships between stimuli. A study by Misyak, Christiansen, and Tomblin (2010) found an association between statistical learning ability and reading-time at the main verb in a sentence containing an object-relative clause (e.g., *the reporter that the senator attacked <u>admitted</u> the error*). Individuals who were better at learning the non-adjacent dependencies in the statistical learning task also processed the long-distance dependency between the head noun and main verb more efficiently in a self-paced reading paradigm. Importantly, the better learners did not show significantly faster reading times when reading the main verb in subject-relative clauses (e.g., *the reporter that attacked the senator <u>admitted</u> the error*).

A similar reading-time effect exists for individuals who are more sensitive to a grammar relying on the learners' ability to track adjacent dependencies (Misyak & Christiansen, 2010). The better an individual was at learning the adjacent dependencies in a statistical learning task the more interference they experienced when processing subject-verb number agreement with conflicting local information (e.g., *the key to the cabinet<u>s was</u> rusty*). This suggests that such learners are hyper-sensitive to adjacent relations even when it was misleading, as all sentences of this type were grammatical. Of note, individual differences in adjacent and non-adjacent statistical learning ability are not correlated with one another (Misyak & Christiansen, 2010).

The individual differences literature on statistical learning further clarifies the relationship between statistical learning and language. Findings which demonstrate that better statistical learning abilities are related to greater language skill validate the idea that statistical learning itself is a contributing factor in language learning and processing, although a direct causal link cannot be inferred due to the correlational nature of these findings. The nuanced literature surrounding the relationship between adjacent statistical learning, non-adjacent statistical learning and language also contributes to the idea that this domain-general process plays an important role in language. It remains to be seen whether the same underlying neural circuitry subserves adjacent and non-adjacent statistical learning, although some recent findings suggest that both can be tracked simultaneously under certain conditions (Vuong, Meyer, & Christiansen, in press).

Individuals with greater experience tracking the types of relationships involved in processing sentences with non-adjacent dependencies should not only show higher performance on language tasks involving such dependencies, they should also show similar performance on tasks that rely on the same types of structure in other domains. This is consistent with other evidence pointing towards the effect that frequency has on processing (e.g., Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). As individuals track the same types of relationships over and over in language, we would expect them to learn the underlying associations between elements that reduce uncertainty if they possess a mechanism for extracting such patterns. Wells et al. (2009) have demonstrated that experience with the reading of relative clause sentences facilitates object-relative clause reading times in adults, demonstrating the importance of experience for language processing, and also providing compelling evidence for the plasticity of entrenchment throughout development. Learners track relationships between linguistic elements over the course of experience, and use the information in these relationships to continuously update their expectations and representations – statistical learning abilities can be thought of as mediating the effect of linguistic experience. Thus, even adults can become better at processing complex linguistic structures once those structures have become entrenched through experience-dependent learning mechanisms, indicating that it is a continuous, lifelong process of learning in language use (see Christiansen & Chater, 2016, for discussion).

The individual differences literature shows that there is variation across individuals in how good they are at picking up regularities given their linguistic experience. These differences highlight the importance of statistical learning in the entrenchment of linguistic structures, and linguistic relationships more generally; increased experience with certain structures leads to more automatic processing of those structures.

5. Statistical learning in models and theories of language learning and processing

Statistical learning is clearly related to some aspects of language learning and processing. Can models and theories of language learning and processing incorporate this mechanism and show that it helps to explain linguistic development?

5.1. Statistical learning leads to entrenched linguistic constructions

Usage-based approaches to language (e.g., Tomasello, 2003; Goldberg, 2003) argue that grammatical knowledge is learned via the chunking/entrenchment of multi-word utterances, rather than relying on innate language-specific knowledge (e.g., Pinker, 1999). Language users have since been shown to rely on such chunks when processing language (see Arnon & Christiansen, submitted, for a review). For example, young children are able to repeat words in highly frequent non-idiomatic chunks more rapidly and accurately than when the same words form lower frequency chunks (Bannard & Matthews, 2008). Adults have also been found to have processing advantage for high-frequency multiword chunks (Arnon & Snider, 2010; Janssen & Barber, 2012), an effect that is modulated by the meaningfulness of the utterance (Jolsvai, McCauley, & Christiansen, 2013). This set of findings indicates the importance of entrenchment to language processing and also highlights the importance of conventionalized form-meaning mappings, supporting construction grammar approaches to language (e.g., Goldberg, 2003). Language users seem to chunk multiple words together in ways that improve processing; these constructions are best understood as entrenched linguistic elements.

How might statistical learning operate as a mechanism for the construction of such chunks? Sensitivity to statistical relationships, like the backward transitional probabilities that infants as young as eight-months are capable of tracking (Pelucchi, Hay, & Saffran, 2009), has been built into certain models attempting to understand how children might form their early lexicon through the construction of these entrenched chunks. The peaks and dips in forward transitional probability have also been identified as potential cues for placing phrasal boundaries when computed over word classes (Thompson & Newport, 2007).

McCauley and Christiansen (2011) have created a model which is capable of tracking the statistical relationships between single words and, based on these relationships, forming chunks. The model is trained on corpora of child-directed speech from the CHILDES database (MacWhinney, 2000), giving it a naturalistic input from which to learn. The model is able to accurately place boundaries between phrases, and also out-performs competing models when attempting to re-produce the utterances of the children in the corpora. In addition, the model parallels child performance in an artificial grammar learning paradigm (Saffran, 2002) when the learning takes place over individual items, rather than classes of items, mirroring its relative performance in the analyses of language production and comprehension, contradicting the findings of Thompson and Newport (2007). This model demonstrates that entrenched units can be formed on the basis of distributional information alone, identifying statistical learning as a mechanism of entrenchment in the contexts of both natural and artificial language.

6. Conclusion

The ability to track and learn probabilistic dependencies between elements seems to be a property of the way that humans learn in multiple domains. Whether the elements are tones (Saffran et al., 1999), syllables (Saffran et al., 1996a; 1996b), word-like units (Gomez, 2002), visual scenes (Fiser & Aslin, 2002), or complex audiovisual stimuli (Mitchel, Christiansen, & Weiss, 2014; van den Bos, Christiansen, & Misyak, 2012), humans are able to learn about the statistical structure underlying their co-occurrence. This evidence points towards statistical learning as a robust, domain-general process (Saffran & Thiessen, 2007), likely implemented in separate modality-specific neural networks relying on similar computational principles (Frost et al., 2015).

The manner in which statistical learning operates, by tracking relational and distributional information for items across space and time leads to the entrenchment of learned relationships. The degree of entrenchment varies between items as a function of frequency (Reali & Christiansen, 2007), meaningfulness (Jolsvai et al., 2013), and predictability (Aslin et al., 1998), and is fundamentally plastic throughout the lifespan (Wells et al., 2009). This general understanding of how statistical learning leads to the construction of units that contain meaning fits well into emergent, experience-based theories about language (i.e., Goldberg, 2003; Bybee, 2006; Elman et al., 1996; Christiansen & Chater, 2016, in press), and identifies it as integral to theories postulating that language learning and processing rely on sensitivity to multiple cues in the input (Christiansen, Allen & Seidenberg, 1998). Highly entrenched items can be stored as chunks, which can become the building blocks of language in development (McCauley & Christiansen, 2011), and which can also affect language processing (Bannard & Matthews, 2008). These entrenched representations are built up over the course of development as a result of statistical learning, allowing higher level linguistic features to be learned.

References

Amso, D., & Davidow, J. (2012). The development of implicit learning from infancy to adulthood: item frequencies, relations, and cognitive flexibility. *Developmental Psychobiology*, *54*(6), 664–73.

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in Psychology*, *3*(August), 324.

Arciuli, J., Torkildsen, J. V. K., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*(July), 747.

Arnon, I. & Christiansen, M.H. (submitted). *Multiword units as building blocks for language*.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436-461

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.

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1425.

Bly, B. M., Carrión, R. E., & Rasch, B. (2009). Domain-specific learning of grammatical structure in musical and phonological sequences. *Memory & Cognition*, *37*(1), 10–20.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review*, *70*(4), 323–348.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

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.

Christiansen, M. H. & Chater, N. (in press). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*.

Christiansen, M.H. & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489–508; 509–58.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.

Conway, C. M. & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus specific representations. *Psychological Science*, *17*, 905–912.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*(4), 597–612.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, *64*(5), 1021–1040.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Franco, A., & Destrebecqz, A. (2012). Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology*, *8*(2), 144–54.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical Learning of Two Artificial Languages Presented Successively: How Conscious? *Frontiers in Psychology*, *2*(September), 1–12.

Frost, R., Armstrong, B. C., Siegelman, N. & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*, 117–125.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Gibson, E. J. (1991). *An odyssey in learning and perception*. Cambridge, MA: MIT Press.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*(6), 249–58.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R.L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

James, W. (1890). The Principles of Psychology, Holt.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PloS One*, *7*(3), e33202.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Jost, E., Conway, C., Purdy, J. D., Walk, A., & Hendricks, M. (in press). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research.*

Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–40.

Kirkham, N. Z., Slemmer, J. A, & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149.

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin*, *126*, 109–137.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk, Vol. II: The database. Mahwah, NJ: LEA.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 1619-1624)*. Austin, TX: Cognitive Science Society.

McCauley, S. M. & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon*, *9*, 419–436.

Miller, G. A. (1958). Free recall of redundant letter strings. *Journal of Experimental Psychology*, *56*(6), 485–491.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology*, 176–185.

Misyak, J. B., & Christiansen, M. H. (2010). When ‘more’in statistical learning means ‘less’ in language: individual differences in predictive processing of adjacent dependencies. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2686–2691).

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Language Learning*, *62*(1), 302–331.

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153.

Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). Statistical-sequential learning in development. *Statistical learning and language acquisition. Berlin: Mouton de Gruyter*.

Mitchel, A. D., Christiansen, M. H. & Weiss, D. J. (2014). Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Frontiers in Psychology*, *5*, 407. doi: 10.3389/fpsyg.2014.00407.

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.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863.

Reber, A. S. (1976). Implicit learning of synthetic languages: the role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 88–94.

Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford University Press: New York.

Reber, A. S. & Allen, R. (2000). Individual differences in implicit learning. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227-248). Philadelphia: John Benjamins.

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27*, 10073–10075.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104.

Saffran, J. R. (2001). The Use of Predictive Dependencies in Language Learning. *Journal of Memory and Language*, *44*(4), 493–515.

Saffran, J. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

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.

Saffran, J. R. & Thiessen E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 68–86). Cambridge: Blackwell.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703.

Schmid, H. J. (2007). Entrenchment, salience, and basic levels. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of cognitive linguistics* (pp. 117-138). Oxford, England: Oxford University Press.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC neuroscience*, *10*(1), 21.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42.

Tomasello, M. (2003). Constructing a language: A usage- based theory of language acquisition. Cambridge: Harvard University Press.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 399.

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*(4), 507–520.

van den Brink, D., Brown, C., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

Vapnik, V. N. (1999). An overview of statistical learning theory. *Neural Networks, IEEE Transactions on*, *10*(5), 988-999.

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (in press). Concurrent learning of adjacent and nonadjacent dependencies. *Language Learning.*

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18,* 414–420.