25

Working Memory and L2 Grammar Development in Children

Paul Leseman and Josje Verhagen

25.1 Introduction

Previous research has shown that individual differences in children's working memory abilities are correlated with their second language (L2) learning outcomes, for both vocabulary and grammar (Cheung 1996; Engel de Abreu & Gathercole, 2012; Kormos & Sáfár, 2008; Service & Kohonen, 1995; Verhagen & Leseman, 2016). While most earlier work has looked into the factors underlying the relationships between working memory and L2 vocabulary, especially regarding the involvement of long-term L2 knowledge (Cheung, 1996; Speciale et al., 2004), far less research has tried to explain the relationships between working memory and L2 grammar. In this chapter, we address this issue. Specifically, we ask why children with better developed working memory abilities generally acquire morphological and syntactic rules in an L2 more readily than children with less-well-developed working memory abilities. In the attempt to answer this question, we explore to what degree several statistical learning frameworks that have been formulated within the field of first language (L1) acquisition can account for the role of working memory in L2 grammatical acquisition. Statistical learning frameworks assign a pivotal role to processes associated with working memory, such as chunking and storing of patterns, to explain the acquisition of linguistic structures by children learning novel or first language forms and structures, but have, to the best of our knowledge, not yet been widely applied to child L2 acquisition.

The organization of this chapter is as follows. First, we describe current concepts of working memory, with particular reference to how working memory is assumed to relate to long-term memory knowledge. We then provide an overview of earlier research on relationships between working memory and L2 grammar. Subsequently, we discuss statistical learning accounts that have been proposed in the field of L1 acquisition, as well as earlier accounts of how working memory processes may be involved in

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

statistical learning. In the core part of our chapter, we address the extent to which these statistical learning accounts can be extended from L1 acquisition to L2 grammar learning, drawing from earlier working memory–based accounts of L2 learning. Finally, in our concluding remarks, we discuss to what extent working memory–based statistical learning accounts can account for the acquisition of L2 grammar, as well as any further steps to be taken to extend their scope from L1 to L2 acquisition.

25.2 Models of Working Memory

Working memory is generally defined as the capacity to hold a small amount of information in a temporary heightened state of activation to make this information available for further processing (Oberauer et al., 2018). One of the most-well-known models of working memory is the tripartite componential working memory model proposed by Baddeley and Hitch (1974), which contains modality-specific short-term memory systems for verbal and visuospatial (and kinesthetic) information, an episodic buffer for temporary multimodal integration and a central executive, which monitors and controls the information flows between the subsystems. After several revisions in the past decades (e.g., Baddeley, 2010), the current version of the model contains a central executive that has no storage capacity itself, but uses the limited capacity short-term memory systems for storage and the episodic buffer for constructing temporary integrated episodic representations. Long-term memory is not considered part of working memory in this model, but the central executive controls the storage and retrieval of processed information in and from long-term memory, and uses knowledge in long term memory for supporting working memory. Specifically, it refreshes or repairs the information in short-term memory to prevent forgetting, and it supports the chunking of information so as to increase the capacity of working memory.

Other working memory models that have been proposed in the literature consider the actual processing of information as part of working memory, with limited capacity temporary storage systems being placed outside of working memory, while working memory proper is equated with limited capacity executive attention. In these models, mainly advocated by Engle and colleagues (Engle, 2002; Unsworth & Engle, 2007; Unsworth & Spillers, 2010), executive attention regulates processing, which is defined broadly, and includes reasoning and pattern matching. A larger role for long term memory has been proposed in models that define activated parts of long term memory as "long-term working memory" (Ericsson & Kintsch, 1995).

In this chapter, we build on the generic model of working memory, as proposed by Cowan (2017). In this model, working memory is seen as a process embedded in long-term memory representations. Most of these representations are in a state of intermediate activation, and a much smaller number is in a state of high activation, usually three or four separate information items, which may also be chunks (Cowan, 2010). In this generic model of working memory, the distinctions between short-term memory, working memory, and long-term memory are not as strict as in the componential model of Baddeley and Hitch, but depend on how the memory system is involved in particular tasks. A limited capacity attention function plays a key executive role in allocating activation, that is, in activating, deactivating, and suppressing memory representations, to make them selectively available for processing (cf. Unsworth & Spillers, 2010). The model is specifically designed to account for the phenomenon that recently activated, but forgotten or suppressed, information is easily reactivated if demanded by the task. This is because this information is assumed to remain in an intermediate state of activation, and thus needs fewer attentional resources to be reactivated. As such, the model can account for the involvement of working memory in complex tasks such as language comprehension and reading, where understanding of a stretch of connected discourse or text requires relating the (propositional interpretation of the) last heard or read sentence to the (propositional interpretations of the) preceding sentences (Ericsson & Kintsch, 1995; Zwaan & Radvansky, 1998).

25.3 Working Memory and L2 Grammar Acquisition

In recent years, accumulating empirical evidence has shown that individual differences in working memory ability across children are correlated with differences in L2 acquisition, for both vocabulary (Cheung, 1996; Masoura & Gathercole, 2005; Messer et al., 2010) and grammar (French & O'Brien, 2008; Kormos & Sáfár, 2008; Service, 1992; Service & Kohonen, 1995; Verhagen et al., 2015; Verhagen & Leseman, 2016). In some of the studies on L2 grammar, relationships between working memory and L2 grammar disappeared once differences in L2 vocabulary were controlled. Service (1992) and Service and Kohonen (1995), for example, investigated the relationship between phonological storage and scores on several grammar tests in Finnish school-aged learners of English. In both studies, phonological storage, as assessed with nonword repetition, predicted children's grammar test scores several years later. However, if vocabulary was included in the analysis, significant relationships were no longer found. Similarly, French (2006) found that effects of phonological storage on English grammar in French child learners of English were mediated by differences in L2 vocabulary knowledge.

Not all studies found that relationships between working memory and L2 grammar were accounted for by L2 vocabulary, however. In a study by French and O'Brien (2008), for example, phonological storage, as assessed with nonword repetition tasks, was a significant predictor of 11-year-old

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

Francophone learners' grammar skills in English, above and beyond English vocabulary. Verhagen et al. (2015) found that verbal memory, assessed through serial word and nonword recall, predicted the production of a number of grammatical structures in Dutch narratives, including subject-verb agreement and verb placement, in Turkish 4-year-old learners of Dutch. The relationship between children's memory scores and grammatical production accuracy remained if differences in Dutch vocabulary were controlled. In another study, Verhagen and Leseman (2016) examined how the working memory components short-term storage and processing, as assessed with a set of different tasks, related to the knowledge of L2 vocabulary and grammar in Turkish child learners of Dutch and a monolingual Dutch comparison group. For grammar, both morphology and syntax were studied. The results showed that verbal short-term storage was significantly associated with vocabulary, while both short-term storage and processing were associated with grammar, for both syntax and morphology alike. There were no differences in the strengths of these relationships between the L2 and L1 learners, suggesting that the same working memory mechanisms are employed for learning vocabulary and grammar in L2 and L1 children. These results indicated, furthermore, that the processing component of working memory is uniquely needed for L2 grammar learning, but not vocabulary, if both vocabulary and grammar are considered simultaneously.

Two points are noteworthy about these earlier studies. First, the children in both Verhagen et al. (2015) and Verhagen and Leseman (2016) were naturalistic L2 learners, who learned the L2 in an immersion setting with little to no explicit instruction. The fact that, in these studies with naturalistic learners, relationships between working memory and grammar remained even if vocabulary was controlled, whereas in studies looking at L2 classroom learners, vocabulary mediated the effects (French, 2006; Service, 1992; Service & Kohonen, 1995), suggests that relationships between working memory and L2 grammar may be specific to children who learn the L2 in uninstructed settings. However, in a study by Engel de Abreu and Gathercole (2012), who looked at trilingual Luxembourgian children learning both the L2 (German) and their L3 (French) at school, significant associations between measures of phonological storage and processing were found. This may indicate that, also for instructed L2 learners, working memory is implicated in the acquisition of L2 grammar.

Second, an important note about the earlier studies on working memory and grammar in child L2 learners pertains to the design of these studies. Without exception, studies were correlational. Therefore, they leave unclear whether the relationships are causal. That is, these studies do not provide insight into the direction of the relationship between working memory and L2 grammar. However, evidence from a number of other studies on L2 children has indicated that the relationships between working memory and language learning may well be reversed such that increased

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

language knowledge supports working memory capacity. Evidence for this idea comes from frequently reported effects of word-likeness and phonotactic probability on tasks in which nonwords are repeated, showing that recall is superior for nonwords that are higher in wordlikeness and phonotactic probability than for nonwords lower in wordlikeness and phonotactic probability in L2 and L1 learners alike (Coady & Evans, 2008; Gathercole, 1995; Messer et al., 2010). For L2 learners, moreover, studies have shown that relationships between measures of working memory, in particular nonword repetition, and vocabulary are typically stronger within than across languages (Lee et al., 2013; Thorn & Gathercole, 1999). Finally, evidence for effects of long-term knowledge on working memory capacity comes from the finding that L2 children typically obtain the highest scores in working memory tasks based on the language they know best (Masoura & Gathercole, 1999; Messer et al., 2010).

A study showing effects of long-term knowledge on working memory in child L2 learners was conducted by Messer and colleagues (2010). These authors investigated serial nonword recall, typically considered a measure of verbal storage, in monolingual Dutch children and sequentially bilingual Turkish 4-year-old learners of Dutch. They administered tasks with nonwords that were composed of phoneme combinations that were high-frequent (i.e., high-probability nonwords) or low-frequent (i.e., low-probability nonwords) in either Dutch or Turkish. The results showed that the Dutch monolingual children obtained the highest scores for recall of Dutch-based nonwords of high-probability, whereas the Turkish-Dutch children obtained the highest scores for recall of Turkish-based high-probability nonwords. Recall of nonwords of low-probability in both Dutch and Turkish was equally low in the Dutch and Turkish-Dutch children. These findings demonstrate that existing language knowledge, in this case knowledge of the phonotactics of both languages, aided children's ability to store the nonwords in working memory. The study also examined how the different types of recall were associated with the monolingual children's Dutch vocabulary and the Turkish-Dutch children's Dutch and Turkish vocabulary. The results revealed unique and significant contributions of both types of recall to the variance in vocabulary in both languages and in both groups. Thus, working memory, as assessed with tasks involving low-probability nonwords, was significantly correlated with children's vocabulary, even when differences in working memory, as assessed with high-probability nonwords, were controlled.

In a study by Messer et al. (2015) the idea that acquired language knowledge facilitates working memory capacity was investigated further longitudinally. Specifically, this study investigated to what extent growth in children's working memory that can be observed as they become older can be explained by growth in language knowledge. To test this, the findings from the 4-year-old children in Messer and colleagues (2010) were extended with assessments from the same children when they were 5 and

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

6 years old. Thus, children's recall of (Dutch-based) high-probability nonwords and recall of (Dutch-based) low-probability nonwords was investigated through ages 4 to 6. Crucially, the authors hypothesized that if improvement in working memory with increasing age was due to growing language knowledge, there should be improvement of high-probability nonword recall (as language knowledge supports the recall of such nonwords and language knowledge typically increases during this period), but no or less improvement of low-probability nonword recall (since increasing language knowledge cannot be used to support the recall of these nonwords). This is exactly what was found: Growth modeling analysis of data from 72 monolingual Dutch children and 69 Turkish child learners of Dutch showed growth in children's recall of high-probability nonwords from the ages 4 to 6, but no growth in recall of low-probability nonwords. These findings held for both groups. Thus, the results of this study indicate that verbal short-term memory growth can be explained by increases in long-term language knowledge (in this case phonotactic knowledge), due to increased exposure to language with age, indicating that not only is there an effect of working memory on language acquisition, as assumed in earlier work, but also vice versa, such that increases in acquired language knowledge result in better-developed working memory abilities.

25.4 Statistical Learning in (Monolingual First) Language Acquisition

Whereas earlier studies into the relationship between working memory and language learning in L2 children have mainly argued that children must be able to hold (phonological) information temporarily active in short-term memory to learn the words and rules in language, accounts formulated within the field of monolingual language acquisition have envisioned a more detailed picture of the complex interplay between perceiving, extracting, and integrating speech information to learn forms, structures, and meaning from language (e.g., Thiessen et al., 2013). Specifically, these studies have tried to uncover how young children learn language from the ambient input by picking up statistical regularities in this input, usually through experimental paradigms using novel language input or artificial languages. The ultimate aim of these studies is to understand the processes underlying the bidirectional relationships between working memory and language learning. Typically, studies have looked at very young children in the earliest phases of language learning, between zero and two years.

In children this young, multiple distributional and frequency cues have been found to support (young) children's detection of language structure (Thiessen et al., 2013), a process that has been referred to as statistical learning. Statistical learning thus refers to learning co-occurrence patterns or distributions based on frequency information in the input. In a seminal study with young infants, Saffran and colleagues (1996) found that infants who were presented with a continuous speech stream of syllables with varying transitional probabilities picked up on these probabilities rather quickly, and were able to accurately distinguish between syllables with high versus low transitional probabilities in a subsequent test. These findings were taken as evidence that children can learn boundaries between combinations of syllables (i.e., "words") from mere frequency information. Other infant studies demonstrated that young children are able to learn co-occurrence relations between elements, even if these are separated by intervening elements, in so-called nonadjacent dependency learning studies. Gómez and Maye (2005) showed, for example, that 15-month-old infants tracked the co-occurrence relationship between structures of the type a-X-b in which a and b were held constant, and X varied. Similar results have been obtained for older children and adults (de Bree et al., 2017; Misyak & Christiansen, 2012; Verhagen & de Bree, 2020).

A central debate about the nature of these mechanisms underlying statistical learning centers around the following question: Does the brain compute probabilities or is statistical learning and, more generally, pattern detection, an emergent property of the interplay between perception and memory? The empirical and computational evidence to date favors the view according to which working memory processes of chunking, attentional biasing and prediction are involved in statistical learning (Dale et al., 2012; Hamrick, 2014; Isbilen et al., 2020; Sherman et al., 2020). According to the Extraction and Integration Framework formulated by Thiessen and colleagues (Erickson & Thiessen, 2015; Thiessen et al., 2013; Thiessen, 2017), for example, 'category' learning requires a memory system that (1) extracts, encodes and stores exemplars (e.g., particular temporal sequences of rising and falling pitch) from input based on conditional statistics, resulting in representations of exemplars which are initially episodic, idiosyncratic and noisy (i.e., also containing features that are less relevant), (2) matches new inputs to existing memory representations and activates the set of those representations that are most similar to the input (a mirroring mechanism) and suppresses or lets decay representations that match the input less well, (3) integrates the new input into the set of best matching representations while changing the internal association parameters of the set to reflect the features that are common across exemplars, after many inputs resulting in a prototype, and, (4) biases subsequent extraction processes to extract units that match the emerging category, thereby enhancing learning efficiency. This memory-based Extraction and Integration framework can explain frequency effects in statistical learning tasks, serial order effects in memory tasks, and non-adjacent dependency learning based on distributional characteristics of the input, that is, the learning of higher order distributional patterns that underlie syntactic knowledge (Erickson & Thiessen, 2015; Thiessen, 2017).

A basic assumption in the L1-based statistical learning studies discussed above is that statistical learning of language at different levels of complexity, from phonemes to syllables and words, from words to morphology and syntax, and to discourse, would require a huge amount of time and exposure if it were not biased by parallel processes to reduce uncertainty, involving attention and memory systems, and additional non-linguistic information (Romberg & Safran, 2010; Thiessen et al., 2013). Therefore, a further assumption is that language learning in naturalistic situations receives additional support from the non-linguistic environment, in particular from the frequency distribution of co-occurrences of words (or phrases and sentences) and particular scenes with potential referents – referred to as cross-situational statistical learning (Romberg & Safran, 2010; Smith et al., 2014). These co-occurrences of words and scenes are relatively unambiguous for the young language learner, often carefully orchestrated and frequently repeated as routines in daily family life (Tamis-LeMonda et al., 2019; Weizman & Snow, 2001), and supported by prosody (Romberg & Safran, 2010) and pragmatic principles such as joint attention (Morales et al. 2000; Mundy & Gomes, 1998; Tomasello, 1988). As such, they form rather salient frequency-based chunks in the non-linguistic visuo-spatial array of objects, actions and events that regulate children's attention to discard irrelevant information, activate relevant memory representations, and become rather consistently (in a statistical sense) associated with linguistic structures.

25.5 Relationships between Statistical Learning and Working Memory

Individual differences in statistical learning ability, the stability of these individual differences, the predictive value of statical learning for language learning, and the relationships between statistical learning and working memory or other domain-general cognitive functions (intelligence, executive function) have received relatively little attention until recently (Siegelman et al., 2017). Based on the extant research on adults, statistical learning is likely not a single domain-general capacity, but rather a set of statistical learning mechanisms that differ by modality (auditive vs. visual), statistical regularity (transitional vs. distributional), contingency (adjacent vs. non-adjacent), and material (verbal vs. non-verbal), that in the complex process of language acquisition may work in parallel (cf. Kidd, 2012). This suggests a multi-componential capacity (Arciuli, 2017; Siegelman et al., 2017) – a conclusion that is supported by the fact that, although separate measures of statistical learning have good test-retest stability, correlations between different measures are generally modest (McCauley et al., 2017; Siegelman & Frost, 2015). Thus, a likely explanation for these modest correlations across statistical learning tasks is that statistical learning involves modality-specific perception, regularity-specific extraction, and modality-specific memory storage, updating and integration (Thiessen et al., 2013). All these subprocesses may require attentional resources, in line with the generic working memory model (Palmer & Mattys, 2016), thus sharing variance caused by individual differences in executive attention ability (Unsworth & Spillers, 2010). However, if the four subprocesses of statistical learning are indeed multiplicatively related, as proposed by Siegelman et al. (2017), it is not surprising that intercorrelations between different statistical learning measures that draw upon different modalities, statistical regularities, contingencies and materials, as well as the associations between statistical learning measures and domain-general working memory measures are moderate at best.

Regarding the relationships between statistical learning and working memory, previous findings have been inconclusive. Siegelman and Frost (2015) examined lower level statistical learning (e.g., word and pattern segmentation) in adults in different modalities with different types of statistical regularities and contingencies. The authors found that different measures of statistical learning were only weakly intercorrelated (despite fair to good test-retest stability), and not related to working memory and other general cognitive functions. Misyak and Christiansen (2012) examined how verbal and visual adjacent and non-adjacent statistical learning were related to verbal short-term memory (forward digit span), verbal working memory (reading span) and several language comprehension measures. Their results showed that verbal working memory correlated moderately strongly (r's in the .40 to .53 range) with both adjacent and non-adjacent statistical learning, while verbal short-term memory significantly correlated with adjacent statistical learning only. Thus, the relationships between measures of statistical learning of language and working memory assessments may differ between low and high level statistical learning, and only the latter may be associated with working memory, possibly due to higher information load and the need to control attention (Palmer & Mattys, 2016). Interestingly, moreover, when regressing the memory and statistical learning scores on tasks of language knowledge and language processing, the authors found unique effects of both adjacent and non-adjacent statistical learning but no direct effects of the verbal short-term and working memory tasks.

For children, only very few studies have looked at the interrelationships between measures of working memory and statistical learning. Typically, these studies examined the correlations between nonword repetition as a measure of verbal short-term memory and non-adjacent dependency learning in artificial language learning experiments. The results revealed, just as for adults (Misyak & Christiansen, 2012), moderate correlations between verbal short-term memory and (higher level) non-adjacent dependency learning (de Bree et al., 2017; Verhagen & de Bree, 2020).

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

25.6 Chunking through Language Learning

Working memory capacity is limited to three or four items, according to the generic working memory model (Cowan, 2017; Oberauer et al., 2018), but can be substantially expanded by creating larger chunks that function as single units in working memory. Chunks can be defined as groups of items with stronger intragroup than intergroup associations (Brady et al., 2009; Cowan, 2010). Language knowledge in the form of chunked information in long-term memory supports short term memory and the creation of new, larger chunks. For example, recalling a string like USAEUNATO is relatively easy compared to a random sequence of the same letters, because long term knowledge helps to recognize the chunks USA, EU, and NATO.

In language acquisition, statistical learning from input and the chunking of information in working memory based on distributional cues (which supports subsequent recognizing of more complex patterns in the input) work in parallel and interactively. Evidence for this comes from a study by Isbilen and colleagues (2020), who trained subjects in an artificial language. In this language, consonant-vowel syllables formed three-syllable nonwords with high within-word and low between-word transitional probabilities, similar to the stimuli used by Saffran et al. (1996). In a serial nonword recall task with more complex six-syllable items, half consisting of combinations of trained three-syllable nonwords with high internal transition probability and the other half of six-syllable control items with the syllables in varying, pseudorandomized order, recall was superior for the nonwords constructed out of syllables with high internal transitional probabilities compared to the nonwords constructed out of equally trained syllables but with low internal transitional probabilities. Similar results were found when natural language (English) syllable statistics were used to create input strings and recall items, revealing an advantage in serial recall for items made of high-probability words based on the frequency of the constituting syllables in natural language, as in the studies on L2-learning children by Messer and colleagues (2010, 2015) described above. These results indicate that statistical learning based on transitional probabilities between phonemes or syllables results in long-term memory representations of chunks (biphone and triphone units, syllables, words, phrases) that, in turn, influence working memory capacity in serial recall of larger units. Because the items were based on natural language, the memory tasks tapped into the subjects' sensitivity to statistical cues in the natural language input they received, thus reflecting individuals' statistical learning abilities in naturalistic language learning situations.

In a study by McCauley and colleagues (2017), participants' abilities to chunk the input were related to the processing of complex sentences. Two types of chunking ability were investigated: (a) phonological chunking ability (defined as the difference in repetition accuracy of nonwords

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

constructed of high vs. low chunklike syllables, as based on natural language statistics) and (b) multiword chunking ability (defined as the difference in recall accuracy of twelve-word strings composed of four subsets of three high vs. low chunklike words, as based on natural language statistics). Both nonword repetition and multiword recall were superior for high chunklike items compared to low chunklike control items, however, showing considerable individual differences. Furthermore, both types of chunking ability were related to the reading of (a) sentences with embedded object-relative clauses and distractions based on phonological similarity of the subject and object and of the two verbs (e.g., "The cook that the crook *consoles controls* the policemen"), and (b) sentences with long-distance subject-verb number agreement with inserted distracting number-marked nouns (e.g., "The key to the cabinet*s* was rusty from many years of disuse"), but in different ways: subjects with higher phonological chunking ability showed less difficulty processing sentences of type (a), while subjects with higher multiword chunking ability showed less difficulty processing sentences of type (b). Interestingly, the two types of chunking ability based on statistical regularities in natural language at different levels were not correlated.

25.7 Construction Grammar and L2 Learning

Chunking in, and through, language learning is a key process according to the usage-based account of grammar learning, also referred to as the "construction grammar" approach (e.g., Ellis, 2002; Kidd, 2012; Tomasello, 2003). This account of language acquisition focuses on the learning and gradual abstraction of "constructions," starting with the smallest meaning-carrying constructions in languages, such as morphemes and simple words, to more complex phrases and abstract syntactic frames (Tomasello, 2003; Wulff & Ellis, 2018). Constructions are regarded as form-meaning mappings, where meaning can be referential, but also functional (as in the passive, which shifts attentional focus to the recipient of an action), and concrete or abstract. The acquisition process is frequency-driven and assumed to involve general mechanisms of cross-situational statistical learning to map forms to meanings (Kidd, 2012). The most frequent constructions in language use are acquired early and become most entrenched, while very infrequent constructions may never be acquired by some language learners, as also suggested by studies on L2 learners who, despite apparent nativelike proficiency, do not show full nativelike mastery upon close scrutiny (Abrahamsson & Hyltenstam, 2009; Ioup et al., 1994).

The idea of chunking in, and through, language has been applied to L2 learning as well. Specifically, researchers have proposed that L2 learners initially store chunked stretches of speech in memory that are analyzed only later on, leading to increasing L2 proficiency (Martin & Ellis, 2012;

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

Speidel, 1993; Speidel & Herreshoff, 1989). Speidel (1993), for example, argued that L2 learners store grammatical constructions in verbal short-term memory, in the same way they store words, that is, as lexical items. In so doing, they build a "storehouse" of constructions in long-term memory from which they can gradually extract patterns to support their spontaneous L2 speech. Similarly, Ellis (1996) argued that most of L2 learning is in fact item learning at different levels of the language system, including grammar (which he termed "sequence learning"). Ellis proposed that L2 learners initially store sequences in a "database" that they later on use to abstract regularities (i.e., grammatical knowledge) from. When storing sequences, chunking plays a major role and the way in which chunks are formed is dependent on long-term knowledge. Specifically, Ellis assumed that the L2 learner first accumulates a sufficient mass of L2 phrases ("sequences") and then uses the same statistical processes of abstraction that the L1 learner uses to discover or "construct" grammatical rules from this collection of sequences (cf. Tomasello, 2003; Ullman, 2001). As an example, he refers to the well-known acquisitional stages that L2 learners of diverse L1 backgrounds go through when acquiring negation in English: no/not + X ("no happy"), before no/not/don't + V ("they not work"; "he don't go"), before analyzed don't ("she doesn't sleep") (e.g., Schumann, 1978). The first two sequences can be seen as co-occurrence patterns that are prevalent in the input and that L2 learners would store as 'sequences' ("I am not happy," "There is no water anymore," "He does not work," et cetera).

Ellis did not go into detail as to which factors underlie the storage and chunking of L2 sequences. Mere frequency cannot be the only source of information that learners use: some constructions in languages are highly frequent, such as articles, verbal suffixes, the placement of negation and adverbs – yet many L2 learners are known to struggle with these constructions also after extensive exposure. Other factors that may facilitate storage and chunking are phonological salience, the extent to which a form carries a clear meaning, prototypicality, and redundancy (Ellis & Collins, 2009). Below we discuss a few of these factors in more detail, pointing to the combined effects of competition of L1 and L2 knowledge, cue salience, and reliability in the input, and the cumulative amount of input.

25.8 Second Language Learning and Working Memory

L2 acquisition (here broadly defined to include successive bilingualism and foreign language learning) differs from L1 acquisition in a number of ways (MacWhinney, 2005, 2012). First, while L1 learning children learn language and learn about the world at the same time, L2 learners usually already have knowledge about the world at the onset of L2 acquisition. Second, L2 learners already have acquired an L1 at least to some degree, which influences the way they process, comprehend, and produce L2 words and

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

sentences. Thus, L1 learners are learning with a still malleable brain that has not yet become specialized and modularized, whereas L2 learners, depending on their age, approach the language learning task with a brain much more dedicated to specific types of input and processes. Finally, a difference between L1 and L2 learners is that L1 learners are usually strongly supported by their caregivers in the family context, whereas L2 learners often have to pick up the L2 from less frequent, more distributed, and less well-organized language input (Jia & Aaronson, 2003; Leseman et al., 2019; Paradis, 2007; Place & Hoff, 2011; Scheele et al., 2010; Unsworth, 2014).

Despite such differences, observations of child and adult L2 learning suggest that L1 and L2 are tightly interwoven in L2 learning, and that many mechanisms are similar (MacWhinney, 2012; Verhagen & Leseman, 2016). Both L1 and L2 learners need to segment speech into syllables, words, and phrases based on low-level statistical learning from input, learn word and sentence meanings by connecting form and meaning in cross-situational statistical learning, and figure out the patterns that govern word combinations in syntactic constructions based on high-level statistical learning, as proposed by the construction grammar approach. Moreover, within L2 learners, there is transfer and interference from L1 to L2, and vice versa. All this calls for a unified theory. Transfer and interference point to competition between L1 and L2 on different levels (e.g., phonology, morphosyntax, lexicon, syntax) and the outcome of this competition may be, at least in part, determined by statistical cue strength in the input (e.g., frequencies of particular word orders, subject-verb agreement cues, case marking, the role of agency; MacWhinney, 2012). This competition can be understood in terms of competing "'resonances" in working memory of chunked L1 and L2 representations of both low and high-level constructions, as we will detail below.

The Unified Competition Model developed by MacWhinney and colleagues (MacWhinney, 2005, 2012; Li & MacWhinney, 2013) assumes a central role for working memory in L2 learning. Specifically, in order to understand form-meaning mappings and L2-L1 mappings, items from both languages need to be in a temporarily activated state. Given that the capacity of working memory is limited, this requires chunking of constructions (Ellis, 2002; Wulf & Ellis, 2018). In online interactive situations such as in naturalistic L2 settings, attentional load may be heavy, as is evidenced by studies showing that even fully competent bilinguals tend to process sentences more slowly than monolinguals. Attentional load is also dependent on the structural characteristics of the L2. For example, in languages with a predominant subject-object-verb sentence structure (German), the processing load increases substantially when several elements are inserted between the subject and the verb. Long-term knowledge of syntactic frames (which can be L1 knowledge; see below), based on chunked exemplars encountered in language use, is needed to support working memory and

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

alleviate attentional load (Ellis, 2002). Chunking explains growing fluency in L2 (Hulstijn, 2002) and is particularly important for grammar learning, for instance, when acquiring complex inflectional morphology (Ellis, 2003). Chunking is also important at the sublexical level in beginning L2 learners, and helps them to identify the phonological composition of words and phrases, which sets the stage for learning more complex constructions (Gupta & MacWhinney, 1997; Hulstijn, 2002).

Working memory – or, more specifically, the interplay between chunked information in long-term memory, attentional mechanisms, and temporarily activated items – is also involved in cross-language influences from the often stronger entrenched L1 on the still largely to be learned L2. Initially, L2 learners rely strongly on their L1 knowledge to process L2 input at the phonological, lexical, semantic, and syntactic level. The working memory mechanism is that of resonance, which is a bidirectional mechanism: cues in the input trigger activation (or "resonance") of the best matching and most entrenched chunks in long term memory at all these levels of language processing, which initially are most likely L1 chunks in L2 learners. This, in turn, biases the attention to and perception of the input cues in a predictive way, in line with the general extraction-integration model described above (Sherman et al., 2020; Thiessen & Erickson, 2013; Thiessen et al., 2013). For example, if the L1 is well-entrenched, as in late L2 learners, the phonological form of an L2 word is likely to be perceived and represented as if it was an L1 word. Conceptual knowledge in L1 is also transferred to L2 to understand word and sentence meanings in L1, and L2 input is initially understood via L1 (called "parasitic use of L2"; MacWhinney, 2012). This is effective to the extent that words in L1 and L2 map to concepts in a highly similar way, but interferes when particular concepts in L1 and L2 do not match. Only when sufficient proficiency in L2 has been built up, thus after sufficient exposure, is L2 connected directly to the conceptual base.

With respect to grammar, long-term knowledge of the L1 initially strongly biases the perception of and attention to cues for sentence processing such as word order, inflections, grammatical morphemes, definite articles, and noun animacy, yielding what has been called a "syntactic accent." This bias can cause difficulties in language comprehension and production, and thus in learning from input. The influence of L1 only gradually disappears but may never be fully absent even in fully competent bilinguals. Strong entrenchment of L1, in this regard, presents the greatest "risk factor" for L2 learners (MacWhinney, 2012) – in addition, of course, to less exposure to L2 as a consequence of the often later age at which L2 learning starts and the inevitable need to distribute exposure time over two or more languages in multilingual situations, as was discussed above.

There is evidence from L2 acquisition research supporting these claims. In L2 acquisition, L2 learners have been observed to prefer adverbial tense marking over verb inflection (Klein & Perdue, 1997; Verhagen, 2009). In

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

general, L2 learners show more difficulty with acquiring morphemes and closed-class grammatical constructions than with open-class constructions (vocabulary, lexicalized phrases). A number of principles can account for this, pointing to perceptual and statistical learning mechanisms involving working memory: cue availability, saliency, and reliability, and (in L1) learned attentional biases based on well-entrenched chunked constructions. Cue availability is determined by the frequency of the cue in language use. Also, the perceptual salience, or detectability, of cues in the language input matters. Less frequent or perceptually less salient linguistic cues are more difficult to pick-up by statistical learning mechanisms. Grammatical morphemes, although occurring frequently, are usually not stressed and often do not match with L1 constructions, so that there is no long-term knowledge that can support perception and chunking of these cues. Thus, in the latter case there is hardly positive transfer from L1 to L2, although this may differ between pairs of languages. In addition, there can be interference if a similar cue has a different function in the two languages (e.g., the determiner-number cues in English vs. Spanish; Li & MacWhinney, 2013). Cue reliability, or the degree of form-meaning contingency, represents the proportion of times a particular cue gives the correct interpretation of all uses of this cue. Cue reliability can be high for some constructions in a particular language, but low for other constructions. Cues can be frequent in a language, but not reliable because of a weak contingency between the cue and its meaning. For example, the highly frequent *-s* in English denotes the plural form but can also denote third person present tense and the possessive relation (Ellis, 2006). Learning case-marking in Russian is easier than in German, despite the fact that the Russian system is more complex, because Russian case-marking cues are more reliable for sentence interpretation (Kempe & MacWhinney, 1998).

Less frequent, less salient, and less reliable cues, and cues that do not match L1 constructions or even interfere with L1 constructions, require substantially more exposure to be learned statistically, which may not be feasible in multilingual situations.

25.9 Conclusion

Earlier L2 research has shown that working memory is implicated in the acquisition of L2 grammar. However, within these studies on the relationships between working memory and L2 grammar learning, no detailed attempts have yet been made to explain *how* working memory may be involved in L2 grammar acquisition. In this chapter, we reviewed working memory-based statistical learning accounts from the language acquisition literature. We also reviewed a number of theoretically related accounts in the L2 literature that assume a pivotal role for processes associated with working memory (e.g., chunking) in L2 grammar learning.

Based on our discussion of these two strands of research, we propose that the relationships between working memory and L2 grammar learning are complex and most likely bidirectional, just like the relationships between working memory and L1 learning (Messer et al., 2010, 2015; Verhagen et al., 2019). Specifically, we propose that working memory enables children to learn constructions at increasing levels of complexity from the surrounding input involving the creation of chunks in long-term memory, while increases in this long-term language knowledge, in turn, enhance working memory capacity, bias attention, and perception, and thereby accelerate further language learning. Departing from the assumptions made in the generic model of working memory (Cowan, 2017), our chapter has attempted to describe how dynamic and interactive working memory processes involved in statistical language learning that have been assumed for monolingual language acquisition may provide a window on the processes involved in L2 grammar learning, at least in naturalistic situations.

Importantly, L2 learners face a number of challenges that L1 learners do not face. Specifically, they are not only likely to encounter difficulties due to the competition between L1 and L2, but also they typically receive less input and less clear language input than L1 learners. These differences likely make language learning for L2 learners more challenging at all levels of the language system, starting with the ability to perceive and chunk phonemes and phoneme clusters in order to segment the speech stream as a prerequisite for word, phrase, construction, and syntax learning. Working memory enables learners to allocate their attentional resources to perceiving and chunking of the input and to memory representations of L1 and L2, and, as such, modulates the competition between L1 and L2, for instance by deactivating or suppressing L1 activation. A speculative idea to be explored in future research is that individual differences in these attentional resources explain individual differences in L2 learning that are still found even if all other factors that are at stake in L2 learning are kept equal.

The ideas we reviewed and proposed in this chapter were specific to L2 learning in naturalistic situations. However, L2 learning may also take place in classroom settings, where it is supported through modified interactions and instructional materials in several ways. Specifically, interfering biases stemming from L1 may be deliberately addressed as part of metalinguistic instruction. Cues in the input that are not salient may be made salient by explicit instruction and corrective feedback, to focus the learners' attention on these cues (and foster "noticing"; cf. Schmidt, 1990). Cues that are infrequent can be deliberately repeated to support the learning process. Negative evidence for certain constructions, though infrequent in natural language, may be brought to the attention of the L2 learner, too (Treffers-Daller & Calude, 2015). In so doing, learning environments can be optimized to foster working memory-based processes at the level of perceiving, extracting, and integrating information, as well as the interplay between these processes, which are needed to learn forms and structures of the L2 and overcome L1 interference.

Our approach in this chapter is compatible with an influential model of language learning as based on two types of domain-general memory and learning mechanisms with distinct neurobiological substrates (Ullman, 2001): declarative and procedural learning (hence, the DP model), the first associated with lexical knowledge and the second with grammatical knowledge. Declarative learning in the DP model relates to what we have called item-based or exemplar learning, which characterizes L1 acquisition at different levels, from phonology to syntax, in young children acquiring their L1 and initial L2 learning until, by chunking, a sufficiently large storehouse is built up for processes of abstraction of categories and rules. Procedural learning in the DP model relates to these processes of abstraction and rule generation. A recent meta-analysis by Hamrick et al. (2018) provides support for the DP model, revealing associations between nonverbal behavioral and brain measures of declarative learning with lexical knowledge and also grammatical knowledge in both 5- to 10-year-old L1 learners and in beginning adult L2 learners, and between nonverbal measures of procedural learning with grammatical knowledge in experienced adult L2 learners (Hamrick et al., 2018). The double association of declarative learning with both lexical and grammatical knowledge in relatively inexperienced language learners (children, beginning adult L2 learners) may reflect, according to the authors, that declarative lexical knowledge gradually, through chunking mechanisms, feeds into procedural learning, as proposed in the construction grammar account (see also Hamrick, 2014). In line with this, we detailed in this chapter how working memory and statistical structures of the language input interact – through extraction of exemplars, chunking, and integration over chunked exemplars in statistical learning – to generate grammatical knowledge, both in L1 and L2. (Note that no studies were found for the meta-analysis of Hamrick et al. [2018] with older L1 learners to test whether, in older children, grammatical knowledge would also be uniquely associated with procedural learning as in experienced adult L2 learners.)

Human working memory as a domain-general limited capacity resource, characteristics of grammars, and how language use in social contexts is governed by these grammars can be seen as constituting together a complex system in which basic properties of each of the three subsystems codetermine the (phylogenetic, cultural-historic, and ontogenetic) development and operation of the other subsystems. Indeed, as is discussed in the chapters by Hawkins and O'Grady in this volume, the grammars of many languages, if not all, reflect the capacity constraints of human working memory and generate in actual language use different types of statistical regularities to be picked up by general-purpose working memory, while prosodic and pragmatic features of language use in social contexts (e.g., the

https://doi.org/10.1017/9781108955638.031 Published online by Cambridge University Press

use of stress, pauses, pointing, gesturing, repeated language routines, cross-situational form-meaning mapping) can be seen as devices to support language learning through working memory-based extraction and integration of statistical information in the input (Thiessen et al., 2013). Indeed, the stimulus is not poor at all and well-attuned to the human language learner.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306.

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B*, *372* (20160058).

Baddeley, A. (2010). Working memory. *Current Biology*, *20*, R136–R140.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). Academic Press.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology General*, *138*, 487–502.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, *32*, 867–873.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairment (SLI). *International Journal of Communication Disorders*, *43*, 1–40.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163–191.

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

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychological Bulletin Review*, *24*, 1158–1170.

Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology*, *8*, 196–209.

De Bree, E. H., Verhagen, J., Kerkhoff, A. O., Doedens, W. J., & Unsworth, S. (2017). Language learning from inconsistent input: Bilingual and monolingual toddlers compared. *Infant and Child Development*, *26*, e1996.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126

Ellis, N. C. (2002). Frequency effects in language processing and acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition*. Blackwell.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Ellis, N. C., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function: Introduction to special issue. *The Modern Language Journal*, *93*, 329–335.

Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, *104*, 974–986.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108.

Ericsson, K. A., & Kintsch, W. (1995). Long term working memory. *Psychological Review*, *102*, 211–245.

French, L. M. (2006). *Phonological working memory and second language acquisition: A developmental study of Francophone children learning English in Quebec*. Edwin Mellen Press.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*, 463–487.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, *23*, 83–94.

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

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267–333.

Hamrick, P. (2014). A role for chunk formation in statistical learning of second language syntax. *Language Learning*, *64*, 247–278.

Hamrick, P., Lum, J. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492.

Hulstijn, J. H.. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193–223.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73–98.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*, e12848.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131–161.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language. Acquisition*, *20*, 543–587.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184.

Klein, W., & Perdue, C. (1997). The basic variety (or: couldn't natural languages be much simpler?). *Second Language Research*, *13*, 301–347.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*, 261–271.

Lee, H. J., Kin, Y. T., & Yim, D. (2013). Non-word repetition in Korean-English children. *International Journal of Speech-Language Pathology*, *15*(4), 375–382.

Leseman, P. P. M., Henrichs, L. F., Blom, E. & Verhagen, J. (2019). Young mono- and bilingual children's exposure to academic language as related to language development and school achievement. In V. Grøver, P. Ucelli, M. Rowe & E. Lieven (Eds.), *Learning through language* (pp. 205–217). Cambridge University Press.

Li, P., & MacWhinney, B. (2013). Competition Model. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3–5). Blackwell Publishing.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.

MacWhinney, B. (2012). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). Routledge.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*, 379–413.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology*, *34*, 383–388.

Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory*, *13*, 422–429.

McCauley, S. M., Isbilen, E. S., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2681–2686).

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, *105*, 306–323

Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., & Leseman, P. P. M. (2015). Growth of verbal short-term memory of nonwords varying in phonotactic probability: A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language*, *84*, 24–36.

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

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- to 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*, 283–298.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*, 469–482.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S. A., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*), 885–958.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology*, *69*, 2390–2401.

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell handbooks of developmental psychology: Blackwell handbook of language development* (pp. 387–405). Blackwell Publishing.

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, *82*(6), 1834–1849.

Romberg, A. S. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*, 906–914.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of mono- and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*, 117–140.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schumann, J. H. (1978). The acculturation model for second-language acquisition. In R.C. Gingras (Ed.), *Second language acquisition and foreign language teaching*. Center for Applied Linguistics.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163–180.

Service, E. (1992). Phonology, working memory, and foreign language learning. *Quarterly Journal of Experimental Psychology*, *45A*, 21–50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*, 155–172.

Sherman, B. E., Graves, K. N., & Turke-Brown, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20.

Siegelman, N., & Frost, R. (2015). Statistical learning as an ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B*, *372*(20160059).

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258.

Speciale, G., Ellis, N., & Bywater, T. (2004). Phonological sequence learning and short term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*, 293–320.

Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language*, *13*, 69–91.

Speidel, G. E., & Herreshoff, M. J. (1989). Imitation and the construction of long utterances. In G. E. Speidel & K. E. Nelson (Eds.), *The many faces of imitation in language learning* (pp. 181–197). Springer-Verlag.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical transactions of the Royal Society B*, *372*, 1711.

Thiessen, E. D., & Erickson, L. C. (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science*, *22*, 239–243.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792–814.

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology Section A*, *52*(2), 303–324.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences*, *10*, 69–88.

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

Treffers-Daller, J., & Calude, A. S. (2015). The role of statistical learning in the acquisition of motion event construal in a second language. *International Journal of Bilingual Education and Bilingualism 18*, 602–623.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105–122.

Unsworth, S. (2005). *Child L2, Adult L2, Child L1: Differences and similarities. A study on the acquisition of direct object scrambling in Dutch*. (Doctoral dissertation, Utrecht University).

Unsworth, S. (2014). Comparing the role of input in bilingual acquisition across domains. In T. Grüter & J. Paradis (Eds.), *Input and experience in bilingual development* (pp. 181–201). John Benjamins.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406.

Verhagen, J. (2009). Finiteness in Dutch as second language (Doctoral dissertation Free University of Amsterdam).

Verhagen, J., Messer, M. H., & Leseman, P. P. M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, *65*, 417–448.

Verhagen, J., Boom, J., Mulder, H., de Bree, E. H., & Leseman, P. P. M. (2019). Reciprocal relationships between nonword repetition and vocabulary during the preschool years. *Developmental Psychology*, *55*, 1125–1137.

Verhagen, J., & de Bree, E. H. (2020). Effects of bilingualism on statistical learning in preschoolers. *Linguistic Approaches to Bilingualism*, 11(5), 611-639.

Verhagen, J. & Leseman, P. P. M. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, *141*, 65–82.

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*, 265–279.

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to SLA. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields*. (pp. 37–56). John Benjamins.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.