

Grammatical difficulties in children with specific language impairment (SLI):

Is learning deficient?

Hsinjen Julie Hsu¹ & Dorothy V M Bishop

University of Oxford

¹Address for correspondence:

Department of Experimental Psychology,

University of Oxford,

South Parks Road,

United Kingdom,

OX1 3UD.

Tel 44 +01865 271334; fax 44 +01865 281255;

email: julie.hsu@psy.ox.ac.uk

Key words: specific language impairment, grammar, statistical learning

Abstract

Theoretical accounts of grammatical limitations in SLI have been polarised between those that postulate problems with domain-specific grammatical knowledge, and those that regard grammatical deficits as downstream consequences of perceptual or memory limitations. Here we consider an alternative view, that grammatical deficits arise when the learning system is biased towards memorisation of exemplars, and is poor at extracting statistical dependencies from the input. We examine evidence that SLI involves deficits in extracting nonadjacent dependencies from input, leading to reliance on rote-learning, and consider how far this may be part of a limitation of procedural learning, or a secondary consequence of memory limitations.

Specific language impairment

The rapidity and ease with which most children learn syntax has frequently been commented on, and has for many years been taken as evidence for an innate, domain-specific language acquisition device (McNeill, 1966). Nevertheless, there are children who are exceptions to this general rule, and who struggle to master the syntax of their native language. When language learning proceeds slowly or imperfectly in a child of otherwise normal abilities, the child is referred to as having a specific language impairment, or SLI. Many children with SLI have particular problems with grammar. This can be demonstrated using language tasks designed to elicit particular constructions. Thus, mastery of verb inflectional endings may be tested with a probe such as: "Tell me something your mum did yesterday", eliciting responses such as, "she comb her hair", or by asking "What does a dentist do?", with the response "he fix my teeth" (Norbury, Bishop, & Briscoe, 2001). These are not the only kinds of grammatical errors seen in expressive language of children with SLI, but they are more typical of SLI than other error types (Lin, 2006). In addition, children with SLI often demonstrate poor understanding of meanings conveyed by syntactic devices, such as word order or inflectional endings. For instance, they may make errors in selecting the correct picture to match a sentence such as "the chicken on the ball is black", (selecting a chicken on a black ball) or "the fish is eaten by the man" (selecting a fish eating a man) (Bishop, 1997).

In his critique of Skinner's (1957) *Verbal Behavior*, Chomsky (1959) argued that language cannot be acquired through associative learning mechanisms, leading to the conclusion that there must be innate grammatical knowledge. It has subsequently been claimed that, because SLI is a heritable disorder in which syntax is selectively impaired, it constitutes evidence for a domain-specific language acquisition device

(van der Lely, 1997). In line with this view, a number of authors have formulated accounts of SLI in terms of impairment or immaturity of an innately specialised language acquisition system (e.g., Clahsen, 1991; Rice et al, 1995; van der Lely, 1997). However, such a view has been challenged by those who propose that language problems in SLI can arise as downstream consequences of more general nonlinguistic deficits. Most research adopting this latter view has focused on the extent to which impairments in auditory perception or short-term memory can account for language difficulties in SLI (see Bishop, 2006, for brief review). More recently, however, there has been a revival of interest in learning mechanisms of language development, with a reconceptualisation of grammar as involving probabilistic knowledge rather than a system of symbolic rules (Edelman & Waterfall, 2007). This is largely prompted by work contesting the poverty-of-stimulus argument (e.g., Billman, 1989; Redington, Chater, & Finch, 1998), and emphasising children's data-mining abilities when confronted with input that is abundant in distributional regularities or patterns (e.g., Gómez & Gerken, 1999; Saffran, 2001).

We shall start by briefly reviewing work on normal language acquisition that considers statistical learning accounts of acquisition of syntax, and then move on to consider how far this conceptualisation of language learning can inform our understanding of grammatical difficulties in children with SLI.

A probabilistic approach to syntax acquisition

Very young talkers do not behave as if they have abstract grammatical rules. Instead, children's early utterances are organized around concrete and particular words and phrases such as *eat* ____ or *draw* ____ (for a review see Tomasello, 2003), and they do not show awareness of the commonality among words belonging to the same syntactic categories (e.g., Verb, Noun). This suggests that children initially store

heard sentences in an exemplar-by-exemplar fashion, without having system-wide syntactic categories or schemes; grammar emerges as statistical generalizations are made over these stored exemplars (Tomasello, 2000). As system-wide abstract syntactic schemes are needed to support production and accurate comprehension of sentences that a child has never heard before, this early exemplar-based learning account provides an interpretation for performance variations often seen in young children. For instance, one child could correctly comprehend the sentence “*The ball is before the duck*” but misinterpret another sentence of exactly the same syntactic structure such as “*The apple is before the car*”. Similar performance variations could also be observed in young children’s speech. As children grow older and receive more language input, detection of statistical regularities embedded in the input gives rise to more abstract (syntactic) patterns or structures. These statistical regularities could take many forms, including frequency of a single unit (e.g. 'she' is far more common than individual names), frequency of co-occurrence (e.g., yesterday he scored a goal), and the transitional probability of one unit given another (e.g., is running; is chased). In addition, type frequency (e.g., frequency of *V-ed*) and token frequency (e.g., frequency of the surface form *talked*) play a critical role in determining the abstractness of the resulting representations (Tomasello, 2003): Storage of individual exemplars is dependent on token frequency, whereas schematized knowledge and the consequent productivity is dependent on type frequency (Ellis, 2002). Detection and development of abstract syntactic patterns is therefore a first step toward context-independent, automatic and error-free sentence comprehension and production.

Pattern learning of grammatical relations: evidence from artificial grammar learning

Statistical learning of grammatical relations has been studied using artificial languages in which statistical relationships such as frequency of co-occurrence between units or transitional probabilities are manipulated to give rise to structural dependencies in the language. Gómez and Gerken (1999) showed that 12-month-old infants can track frequency of co-occurrence to learn the orderings of words in sequences. They exposed infants to one of two artificial languages. Some transitions between word categories occurred in Language 1 but never occurred in Language 2, and vice versa. After brief exposure to a subset of strings in their training language, infants were tested with novel strings (i.e., strings they had never encountered during training) from both languages. They found that infants listened longer to novel strings from their training language than to strings from the other language, regardless of which language they heard during training. Because infants were never tested on the exact strings encountered during training, it could be concluded that learning was not restricted to memory for particular strings, but rather generalised to novel strings with familiar co-occurrence patterns.

A key difficulty posed by grammar, however, is that it does not just involve learning dependencies between adjacent words. There are also long-distance dependencies. For instance, in "the chicken on the ball is black", it is the chicken, rather than the ball, that is black. Saffran (2001) examined whether statistical learning mechanisms can succeed in mastering more complex structures that are not tied to the surface properties of the input, such as hierarchical phrase structures. For instance, the words in "*The space probe sent back photographs of Mars*" fall into particular groupings (*The (space probe) (sent back (photographs of Mars))*) rather than random grouping (e.g., *(The space) (probe sent back) (photographs of) (Mars)*). To examine if statistical learning could extract hierarchical phrase structure, Saffran compared

learning of two artificial languages, one containing predictive dependencies between words and the other lacking predictive dependencies. The predictive dependencies were defined as relatively higher transitional probabilities between items. In the predictive language the presence of a word token in a category was always preceded by an occurrence of a word belonging to a different word category. In contrast, in the non-predictive language the relationships between categories were variable. Category membership in the languages can be learned by paying attention to the distribution of words. This is similar to the case that speakers of natural languages can infer the category memberships of novel words from surrounding words, e.g. a word preceded by “the” is very likely to be a noun. The training involved a 30-minute session on each of two consecutive days for adult participants and a 21- to 28-minute session for children participants. Both adult and child learners exposed to the language containing predictive dependencies performed better in detecting phrasal units than learners exposed to the language lacking predictive dependencies. Of course, the success in learning does not necessarily entail that the acquired knowledge was hierarchical in nature; nevertheless, the study demonstrates that this type of learning can extend to relationships between word categories that are hierarchically organized, and is not restricted to learning relationships between individual word items, or relationships that can only be sequentially characterized.

Distributional regularities have been shown to be useful in learning other aspects of grammatical relations. These include building syntactic categories from distributional cues in speech (e.g., Gómez & Lakusta, 2004), and learning dependencies between syntactic categories (e.g., Gerken, Wilson, & Lewis, 2005).

Artificial grammar learning has also been used to explore the conditions that promote learning of statistical structure. In order to correctly mark grammatical

agreement one typically has to ignore considerable variation in intervening elements - typically open class words that have a large set size. Consider, for instance, the variable material that can intervene between two grammatical elements that need to be marked for agreement in English structures such as progressives (e.g., the cat is eating), perfectives (He has played the computer game for hours), plurals (The divers on the boat are excited), and third person singular -s (She jumps). Gómez (2002) considered how variability of intervening elements affected learning of nonadjacent dependencies, using artificial 3-word strings, A-X-B, where A and B were always the same, and X was represented by a set of 3, 12, or 24 words. She exposed 18-month-old infants to three-word strings composed of nonsense words such as *pel-chila-tud*, *vot-chila-jic*, in which the occurrence of the first word always predicted the identity of the third word. During a subsequent test phase, the infants' listening time was measured to "grammatical" strings conforming to the dependencies between A and B that had been heard during training and other, "ungrammatical" strings that violated these dependencies. Gómez found that infants discriminated between grammatical and ungrammatical strings only in the high variability condition (24 words). She concluded that listeners seek out statistical regularity in the input: if there are strong dependencies between adjacent elements (as with the small set size condition), the infant will focus on these, and less attention is paid to nonadjacent dependencies. However, when the set of intervening words is large, there is little predictability between adjacent words, and therefore the more stable relationship between nonadjacent elements becomes more salient and thus easy to detect. Note that infants in this study demonstrated sensitivity to statistical regularities in the input in the absence of any corrective feedback. In the same report, Gómez demonstrated similar results with adults, who were required to give grammaticality judgements on word strings

after exposure to novel word sequences: once again, detection of nonadjacent dependencies was best when the transitional probabilities between adjacent items were low. Similar findings were replicated by Onnis, Christiansen, Chater, and Gómez (2003) and Onnis, Monaghan, Christiansen, and Chater (2004).

Specific language impairment: a problem in statistical learning?

As noted above, traditional accounts of grammatical problems in SLI have been polarised between those in the Chomskyan tradition that postulate deficit or delayed maturation of domain-specific grammatical knowledge, and those that regard language impairment as a downstream consequence of more domain-general problems with perception or memory. The novel perspective on grammar-learning provided by those working on statistical learning suggests that we should look more directly at the process of extracting abstract knowledge from statistical regularities in the input, as a possible source of problems in SLI. We will consider four questions relevant to this issue:

- 1) Is there evidence that children with SLI learn language by exemplars rather than abstracting rules?
- 2) How do people with SLI perform on artificial grammar-learning tasks; in particular, do they show the normal sensitivity to transitional probabilities between adjacent and non-adjacent items?
- 3) Insofar as there is evidence of deficient statistical learning in children with SLI, should this be conceptualised as a deficit in procedural learning?
- 4) Could poor performance on statistical learning tasks be due to limitations of perception or short-term memory?

1. Evidence for exemplar-based learning by children with SLI

Gopnik and Crago (1991) were among the first to suggest that children with SLI engaged in exemplar-based learning. They argued that regular and irregular past tenses were treated the same: the child learned the whole inflected form and did not show awareness that the –ed ending was common to the regulars. Goad and Rebellati (1994) provided evidence that even when children appeared to be aware of a rule, in this case plural –s, they did not internalise it, but applied it effortfully by remembering the explicit rule 'add –s'. These authors were working in a Chomskyan framework, and so interpreted their results as indicating a lack of innate grammatical rules, but the findings could be reconceptualised as evidence for a failure of statistical learning mechanisms.

A tendency to use rote-learned forms has also been reported in studies of SLI focusing on learning syntactic structures of new words. Jones and Conti-Ramsden (1997) compared the verb use of three children with SLI and their younger siblings matched on MLU. The children with SLI tended to use lexical verbs in a narrower range of forms than their younger siblings, although their verb lexicons were not dissimilar in size. Interestingly, overlap of utterances with those of the caregiver was greater for children with SLI than for their siblings, suggesting rote learning of syntactic forms. Similar results were reported by Stokes and Fletcher (2000) who investigated spontaneous speech of Cantonese-speaking children with SLI. They found that these children's use of aspect markers was far less productive than language-matched typically-developing children. Furthermore, Skipp, Windfuhr, and Conti-Ramsden (2002) reported context-dependent use of newly acquired noun phrases in children with SLI. In their study, a group of 35-month-old children with SLI learned four nouns presented in one of four argument structures: (1) neither agent

nor patient (e.g., Look- Gabber!), (2) agent only (e.g., The Mogo is pushing), (3) patient only (kissing the Neffy) or (4) both agent and patient (Minnie is washing the Toma). They found that typically-developing children used the noun words in spontaneous utterances equally in all four argument structures, regardless of the argument structures in which the words were presented during training. However, children with SLI demonstrated greater input dependence in terms of the type of arguments they used in spontaneous utterances after training.

Finally, more indirect evidence for difficulties in learning grammatical abstractions comes from consideration of children's performance on grammatical comprehension tasks, where there is no demand for syntactic creativity. At school age most children with SLI are able to respond correctly to many syntactically simple sentences in a multiple choice context. They do not behave as if they have no knowledge of how grammar conveys meaning: rather, they perform inconsistently on more complex constructions such as passives, comparatives, or spatial prepositions (Bishop, 1982). This was demonstrated in an intervention study by Bishop, Adams, and Rosen (2006), who showed that, even after daily training with a particular construction type, fluent automatic comprehension was not achieved, although performance was well above chance. This performance pattern suggests that these children may be poor statistical learners who have not reached a stage at which sentence comprehension is supported by more abstract syntactic patterns required for consistent and accurate performance.

2. Artificial grammar learning in language-impaired individuals

Learning grammar involves learning both adjacent and nonadjacent relationships. For instance, in English serial word order is important to language comprehension as changing the order of noun phrases in a simple Noun-Verb-Noun

sentence (e.g., The boy kicked the girl vs The girl kicked the boy) would result in meaning differences. Plante, Gómez, and Gerken (2002) examined statistical learning of sequential word order in adults with and without language/learning disabilities. Participants first listened to a set of sentence strings conforming to the word order constraints in a finite-state grammar and then provided grammaticality judgments on novel strings. With only 5 min exposure to the language, typically-developing adults were able to exceed chance performance, whereas adults with language/learning disabilities showed chance level performance. Because the speed of presentation of the training stimuli was controlled to be approximately 1.32 words/second, poor performance in the adults with language/learning disabilities, as argued by the Plante et al., was unlikely due to difficulties in processing rapidly presented acoustic information.

Learning of nonadjacent relationships is of particular relevance to the grammatical morpheme deficits that are seen in young typically-developing children and which also constitute one of the most salient features in children with SLI. Before full mastery of verb inflections, young English-speaking children go through a phase where they are inconsistent in use of verb inflections, and produce utterances such as "John go there", while also producing correctly inflected versions of the same verb, e.g., "He goes there". Wexler (1994) argued that this represents a stage where children fail to understand that finite verb forms (forms involving tense and agreement) are obligatory in main clauses, and so treat marking of finiteness as optional. This account has been extended to explain deficits in children with SLI by arguing that they have an unusually prolonged optional infinitive stage (Rice, Wexler, & Cleave, 1995). However, recent empirical studies on pattern learning of remote dependencies in typically-developing individuals provide an alternative account.

Grunow, Spaulding, Gómez, and Plante (2006) adopted the task developed by Gómez (2002) to investigate learning of nonadjacent dependencies in college students with and without language-based learning disabilities. Participants in the study listened to “sentences” composed of three nonsense words in one of two variability conditions: set size of the middle words was either 12 words or 24 words. Adults with normal language were able to learn and generalize the nonadjacent dependencies when variability was high (set size 24). In contrast, adults with language-based learning disabilities did not perform above chance under either variability condition. Grunow et al. concluded that the adults with language-based learning disabilities have poor sensitivity to statistical information in speech input. Unfortunately, the results were not watertight, because the sample size was small ($N = 11$ per condition) and there were no significant group differences in learning nonadjacent dependencies, for either trained items or generalisation items. The conclusion therefore hinged just on the contrasting patterns of mastery across groups and conditions. However, confidence in the conclusions is boosted by findings from another study adopting a similar procedure in adolescents with and without language impairment (Hsu, Tomblin, & Christiansen, 2008). Again, high variability facilitated learning of nonadjacent dependencies in typically-developing adolescents but not adolescents with language impairments. When the participants’ language skills were taken into account, a significant, albeit modest, correlation ($r = .35$) in the high variability condition was found.

Hsu et al. further explored individual differences in this kind of learning by considering the number of participants who reached 100% accuracy in learning at least one nonadjacent pair. Because the test items in their study were not novel strings but strings heard during training, the results would provide further information about

exemplar-based learning. Of particular interest is the children's performance in the variability condition where there were only 2 possible intervening words. In this set size =2 condition, the number of different sentence strings is much lower (i.e., 6) and the frequency of occurrence of individual sentence is much higher (token frequency =72 for each sentence) than that in the other two variability conditions where there were 36 different sentences and each has a token frequency of 12 in the set size = 12 condition, and 72 different sentences and each has a token frequency of only 6 in the set size = 24 condition. If there was a sign of exemplar-based learning, one would most likely to see such learning in the condition where token frequency is high. The results showed that the proportion of typically-developing participants who mastered at least one pair was highest in the high variability condition, as expected. However, for the language-impaired group, 15% mastered at least one pair from the high variability condition versus 25% in the other two variability conditions., An exemplar-based learning account might explain the results. High token frequency could potentially facilitate storage of individual sentence exemplars, although the goal of the task was to train the participants to learn nonadjacent dependencies. In effect, language-impaired participants performed better in the condition where token frequency was highest, suggesting that at least some of these individuals were just memorizing individual strings (i.e., rote learning). As a result, the relatively higher token frequency in the low variability condition became the most facilitative learning condition for these individuals. Again, caution is needed in interpreting these pattern differences because of limited statistical power, but the observed patterns reveal that some of the language-impaired adolescents might have learned the materials in a way different from the typically-developing adolescents, but consistent with literature reviewed above suggesting exemplar-based language-learning in SLI.

3. The Procedural Deficit Hypothesis (PDH) of SLI

Recently, Ullman and Pierpont (2005) put forward a hypothesis of SLI as a procedural learning deficit. This Procedural Deficit Hypothesis (PDH) is an extension of Ullman's Declarative-Procedural (DP) model of language (Ullman, 2001) which postulates a distinct procedural learning system composed of interconnected brain structures, the most important of which appear to be the frontal/basal-ganglia circuits (for a review see Ullman & Pierpont, 2005). According to Ullman (2001), the procedural memory system is important for learning of syntax and phonology, and contrasts with a declarative learning system that is involved in acquisition of vocabulary and more general semantic knowledge.

Historically, the concept of procedural learning was heavily influenced by early studies of neuropsychological patients, especially those with amnesia. Early studies focused on motor skill acquisition, with a striking demonstration that the famous patient HM, who was densely amnesic after bilateral hippocampectomy, could master new motor learning, even though he could not remember doing the task (Corkin, 1968). In a review, Squire (2004) noted that such early observations of preserved motor learning in amnesia were soon supplemented with accounts of spared perceptual learning, and visual and orthographic fragment completion (identifying pictures or words on the basis of partial cues). Of particular interest here were findings that people with amnesia could perform well on tasks involving extraction of prototypes and categories from probabilistic information, and on artificial grammar learning. Does this mean, then, that a single procedural system involved in all aspects of implicit learning (see Figure 1)? Squire (2004) argued against this idea, and proposed instead that there are multiple non-declarative learning systems. There is evidence, for instance, that learning artificial language might not draw on the same

brain structures as skill learning, as patients with basal ganglia dysfunction can accomplish artificial grammar learning (Reber & Squire 1999; Witt, Nühsman, & Deuschl, 2002). With regard to SLI, Ullman and Pierpont (2005) argued that the deficit was in a procedural memory system that controlled learning and control of motor skills such as typing or riding a bicycle, and they drew attention to the motor deficits commonly found in SLI as evidence for their account. Their view of procedural learning is broader than that of Squire (2004), who identified 'procedural learning' with mastery of motor skills and habits, and they postulated a deficit in extracting patterns from verbal material, which also extends to include all types of sequence-learning, "serial or abstract, or sensori-motor or cognitive". In terms of Figure 1, then, the proposed deficit in the PDH encompasses the region encircled by the dotted line, possibly also extending to cover additional aspects of statistical learning.

The PDH generates two types of prediction. First, it predicts that children with SLI will show impairments in implicit sequence learning tasks, whether verbal or non-verbal. Second, the PDH implies that the nature of the language learning problem in SLI has to do with implicit learning of underlying structure from statistical features of the input, rather than rule extraction.

Empirical findings in line with the first prediction came from studies using nonverbal materials to test procedural learning in SLI. Using a serial reaction time task, Tomblin, Mainela-Arnold, and Zhang (2007) tested a group of adolescents with SLI who implicitly learned a ten-item repeating sequence by tracking a stimulus that moved among four spatial locations on a computer screen. Participants pressed one of four buttons on a response panel that matched the location where the visual stimulus appeared. Implicit knowledge of the sequence was revealed by faster reaction times

when subsequently tested on the repeating sequence vs a random sequence. The adolescents with SLI showed a slower learning rate than typically-developing controls. Poor procedural learning was observed by Lum, Gelec, and Conti-Ramsden (2009) who used a similar serial reaction time task and found poor procedural learning of sequence in a group of younger children with SLI compared with age-matched typically-developing children. The same children were unimpaired relative to typically-developing children on a declarative learning task provided non-verbal materials were used. The relationship between grammatical abilities and procedural learning measured was demonstrated in a further analysis by Tomblin et al. who reclassified their participants into groups based on their grammar and vocabulary abilities and found a slower learning rate in the SLI group only when the language impairment was defined in terms of grammatical impairments, suggesting that the same learning mechanisms underpinning motor learning of sequential patterns might also be involved in learning grammar.

Turning to consider the nature of the mechanism of implicit language learning, Perruchet and Pacton (2006) recently compared the ways in which the terms statistical and implicit learning had been used. In their article, they did not consider neuropsychological bases of different implicit memory systems. Instead, implicit learning was used to refer to the general ability to learn sequential patterns when there is no conscious intent to extract this sequential information. The authors argued that although implicit learning and statistical learning have much in common, researchers in these two fields tend to emphasise different learning processes. In the implicit learning literature the focus has been on the way in which grammaticality judgments in artificial language learning tasks rely on knowledge of fragments of strings or chunks (Knowlton & Squire, 1996), and the same process could explain reductions in

response time in serial reaction time tasks (Jiménez, 2008). In contrast, in most artificial grammar research the emphasis has been on how humans learn general statistical information. There is, however, a confound between predictability of stochastic patterns and chunk strength, and grammatical test strings will typically have greater chunk strength than ungrammatical strings.

Recent imaging studies have suggested that both chunk strength and statistical properties of input are used in implicit learning. Liberman and colleagues (2004) used event-related functional magnetic resonance imaging (fMRI) to identify neural regions involved in artificial grammar learning. They controlled for the chunk strength effect, by ensuring that the average chunk strength of grammatical items was equivalent to the average chunk strength of ungrammatical items. They found that activation in the right caudate was associated with pattern learning, whereas medial temporal lobe activations were associated with chunk strength. These findings suggest that both chunk formation and statistical learning of patterns are involved in artificial grammar learning. Interestingly, imaging studies have also provided converging evidence that both striatum (Poldrack et al., 2001; Raucher et al., 1995) and medial temporal lobe (Curran, 1997; Schend et al., 2003) are recruited during serial reaction time tasks.

At this point we may ask whether the PDH hypothesis is the same as a statistical learning account of SLI. Although there is considerable overlap in these viewpoints, they are not identical. The PDH adopts a more polarised view whereby different aspects of language learning are mediated by declarative and procedural systems, which have separate neurobiological bases. Thus the declarative system is seen as handling not only learning of vocabulary and semantic information, but also other arbitrary linguistic information that is not predictable from hypothesised rules,

including storage of irregular past forms (Pinker & Ullman, 2002). In contrast, research in the field of statistical learning has converged on a conclusion that statistical learning mechanisms can achieve grammatical acquisition as well as other aspects of language, such as learning arbitrary associations between sounds and referent (Smith & Yu, 2008) and generalizing names for solid objects by shape in learning vocabulary (Samuelson, 2002). Another example that demonstrates the differences between the two accounts is English past tense morphology. Ullman and Pierpoint (2005) draw a clear line between memory systems underpinning acquisition of irregular past tense (declarative learning) and regular past tense (procedural learning), citing in support disproportionately poorer performance by children with SLI in regular past tense compared with irregular past tense (e.g., van der Lely & Ullman, 2001). However, a recent study has queried this conclusion, showing that when input frequency is controlled, children with SLI making as many errors in marking past tense for regular verbs than irregular verbs (Serratrice, Conti-Ramsden, & Joseph, 2003). In addition, there is evidence that stochastic regularities in the input are sufficient to achieve acquisition of both regular and irregular past tense (Albright & Hayes, 2003).

4. Impact of limitations in perception or short-term memory on statistical learning

Demonstration of poor statistical learning does not necessarily indicate that this is the core deficit in SLI. It could be that extraction of statistical structure from language input is poor because the incoming information is not adequately perceived. Suppose, for instance, that a child had difficulty in discriminating between all the phonemes in the native language. This could have two adverse consequences for statistical language learning. First, some key regularities in the input may be missed:

for instance, it has been noted that in English, past tense and plural endings are brief and may be hard to perceive (Leonard, 1989). This could make it harder to extract regularities involving these endings. Second, if minimal pairs of words are perceived but not adequately distinguished, then the number of perceived types for a given number of tokens, X , in a structure such as $A-X-B$ will be reduced. As noted above, this would make it harder to extract the $A-B$ regularity. Although we cannot rule out this kind of mechanism, it does not seem a sufficient explanation for poor language learning in SLI. First, the syntactic problems of children with SLI are markedly worse than for children with mild-to-moderate hearing loss (Norbury et al., 2001). Second, poor comprehension of grammatical contrasts by children with SLI was found in a training study by Bishop and colleagues (2006), even in optimal listening conditions with simple, pictured vocabulary and short sentences. Furthermore, acoustic modification of the spoken sentences to lengthen and amplify brief and non-salient portions of the speech signal had no beneficial effect. Finally, this account could not explain deficits in statistical learning in nonverbal motor tasks (Tomblin et al, 2007).

Limitations of short-term memory could, however, potentially be implicated in weak statistical learning. Abstract syntactic patterns are built over individual exemplars that are stored in the first place. In order to detect a pattern in input such as $A-X-B$, $A-Y-B$, $A-Z-B$, one must be able to retain at least some three-element sequences in memory. If memory span is only two items long, the pattern will not be detected. Suppose a child with SLI retains only 50% of 3-element utterances, whereas a typically-developing child retains 90% of the same material. This would mean that children with SLI require more exposure in order to memorize sufficient amount of materials before more abstract patterns could emerge, and so language will be more

dependent on memorising of chunks of words. Given that deficits in working memory in SLI are well attested in the literature (for a review see Coady and Evans, 2008), it is possible that the seeming over-reliance on rote learning is a result of limited memory capacity. This suggests that rather than a deficit in statistical generalisations, the observed poor performance in statistical learning could reflect a fundamental difficulty in reaching a “critical mass” (Marchman & Bates, 1994). This interpretation makes testable predictions. Consider the nonadjacent dependency task, for instance. If children with SLI have adequate statistical learning abilities, but their learning is hampered by deficient working memory, they should be slower than other children at reaching an adequate level of performance in grammatical judgment of *previously-heard* sentences, but once that level is reached, they should show normal performance with novel sentences. On the other hand, if working memory does *not* explain poor statistical learning in SLI, then children with SLI should reach an adequate level of grammaticality judgment of sentences they heard during training (exemplar-based), but accurate judgment of novel sentences (generalisation) would still be challenging.

Corpus analyses provide another possible way to compare these accounts. As high token frequency would facilitate exemplar learning, type frequency is critical for generalisation. It is interesting to explore adult language input and test whether children with SLI at a young age produce utterances of relatively high token frequency to a similar level than typically-developing children and, more importantly, do so for a protracted period. On the other hand, a syntactic construction that is high on type frequency and low on token frequency should pose more challenges for children with SLI than a syntactic construction of a similar type frequency but a relatively higher token frequency.

Concluding comments

Recent studies on statistical learning have shown that domain-general statistical learning mechanisms can learn structural patterns. This provides the basis for developing an alternative account of grammatical deficits in children with SLI. The extent to which deficits in statistical learning could supplement extant theories, such as deficits in working memory, in the literature of SLI requires further empirical examination. Work on this topic is of applied as well as theoretical interest. We are gaining increasing knowledge of the conditions that facilitate statistical learning; and this line of research can potentially provide useful information for future development of intervention programmes.

References

- Albright & Hayes (2003). Rules versus analogies in the English past tenses: A computational/empirical study. *Cognition*, *90*, 119 – 161.
- Billman, D. (1989). Systems of correlations in rule and category learning: use of structured input in learning syntactic categories. *Language and Cognitive Processes*, *4*, 127-155.
- Bishop, D. V. M. (1997). *Uncommon Understanding: Development and Disorders of Language Comprehension in Children*. Hove: Psychology Press.
- Bishop, D. V. M. (2006). Developmental cognitive genetics: How psychology can inform genetics and vice versa. *Quarterly Journal of Experimental Psychology*, *59*, 1153-1168.
- Bishop, D. V. M. (1982). Comprehension of spoken, written, and signed sentences in childhood language disorders. *Journal of Child Psychology and Psychiatry*, *23*, 1-20.
- Bishop, D. V. M., Adams, C. V., & Rosen, S. (2006). Resistance of grammatical impairment to computerized comprehension training in children with specific and non-specific language impairments. *International Journal of Language and Communication Disorders*, *41*, 19-40.
- Chomsky, N. (1959). A Review of B. F. Skinner's *Verbal Behavior*. *Language*, *35*, 26-58.
- Clahsen, H. (1991). *Child language and developmental dysphasia: Linguistic studies in the acquisition of German*. Amsterdam: John Benjamins Publishing Co.
- Coady, J. A., Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, *43*, 1-40.

- Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, 6, 255-265.
- Curran, T. (1997). Higher order associative learning in amnesia: evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, 9, 522–533.
- Edelman, S., & Waterfall, H. (2007). Behavioral and computational aspects of language and its acquisition. *Physics of Life Reviews*, 4, 253-277.
- Ellis, N. C. (2002). Frequency effects in language processing – a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Gerken, L., Wilson, R., Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249-268.
- Goad, H., & Rebellati, C. (1994). Pluralization in familial language impairment. In J. Matthews (Ed.), *Linguistic Aspects of Familial Language Impairment, Special Issue of the McGill Working Papers in Linguistics, vol 10*. (pp. 24-40).
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads specific and abstract knowledge. *Cognition*, 70, 109-135.
- Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-monthold infants. *Developmental Science*, 7, 567-580.
- Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition*, 39, 1-50.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39, 158-170.

- Hsu, H., Tomblin, J. B., Christiansen, M. H. (2008). The effect of variability in learning nonadjacent dependencies in typically-developing individuals and individuals with language impairments. In A. Owen (Chair), *The role of input variability on language acquisition and use*. Symposium presented at the XI International Congress for the Study of Child Language (IASCL): Edinburgh, Scotland.
- Jiménez, L. (2008). Taking patterns for chunks: is there any evidence of chunk learning in continuous serial reaction-time tasks? *Psychological Research*, 72, 387-396.
- Jones, M. & Conti-Ramsden, G. (1997). A comparison of verb use in children with specific language impairment and their younger siblings. *Journal of Speech Language and Hearing Research*, 40, 1298–1313.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169-181.
- Leonard, L. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, 10, 179-202.
- Liberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design, *Journal of Cognitive Neuroscience*, 16, 427-438.
- Lin, Y.A. (2006). Against the deficit in computational grammatical complexity hypothesis: a corpus-based study. *Concentric: Studies in Linguistics*, 32, 59-70.
- Lum, J. A. G., Gelgec, C., & Conti-Ramsden, G. (2009). Procedural and declarative memory in children with and without specific language impairment.

- International Journal of Language and Communication Disorders*, 44, 1-19.
- Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development - a test of the critical mass hypothesis. *Journal of Child Language*, 21, 339-366.
- McNeill, D. (1966). Developmental psycholinguistics. In F. Smith & G. Miller (Eds.), *The genesis of language* (pp. 15-84). Cambridge, MA: MIT Press.
- Norbury, C. F., Bishop, D. V. M., & Briscoe, J. (2001). Production of English finite verb morphology: A comparison of SLI and mild-moderate hearing impairment. *Journal of Speech, Language and Hearing Research*, 44, 165-178.
- Onnis, L., Christiansen, M. H., Chater, N., & Gomez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886-891). Mahwah: Lawrence Erlbaum Associates Publishers.
- Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233-238.
- Pinker, S., & Ullman, M. T. (2002). The past and future of past tense. *Trends in Cognitive Sciences*, 6, 456 – 463.
- Plante, E., Gómez, R. L., & Gerken, L. (2002). Sensitivity to word order cues by

- normal and language/learning disabled adults. *Journal of Communication Disorders*, 35, 453-462.
- Poldrack, R. A., Clark, J., Pare´-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414, 546–550.
- Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., Fischman, A. J., & Kosslyn, S. M. (1995). A PET investigation of implicit and explicit sequence learning. *Human Brain Mapping*, 3, 271–286.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425–469.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson’s disease. *Behavioral Neuroscience*, 113, 235-242.
- Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*, 38, 850-863.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.
- Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, 37, 1013–1025.
- Serratrice, L., Conti-Ramsden, K. & Joseph, G. (2003). The acquisition of past tense in pre-school children with Specific Language Impairment and unaffected controls: regular and irregular forms. *Linguistics*, 41, 321-349.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

- Skipp, A., Windfuhr, K., & Conti-Ramsden, G. (2002). Children's grammatical categories of verb and noun: a comparative look at children with specific language impairment (SLI) and normal language (NL). *International Journal of Language and Communication Disorders*, 37, 253–271.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82, 171-177.
- Stokes, S. F. and Fletcher, P. (2000). Lexical diversity and productivity in Cantonese-speaking children with specific language impairment. *International Journal of Language and Communication Disorders*, 35, 527–541.
- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4, 156-163.
- (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Tomblin, J. B., Mainela-Arnold, E., Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, 3, 269-293.
- Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of psycholinguistic research*, 30, 37-69.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41, 399-433.
- van der Lely, H. K. J. (1997). Language and cognitive development in a grammatical SLI boy: modularity and innateness. *Journal of Neurolinguistics*, 10, 75-107.
- van der Lely, H, K, J., & Ullman, M. (2001). Past tense morphology in specific language impaired and normally developing children. *Language and Cognitive Processes*, 16, 177-217.

- Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D.Lightfoot and N. Hornstein (eds), *Verb Movement*. New York: Cambridge University Press, pp: 305-350.
- Witt, K., Nühsma, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia*, *40*, 1534-1540.

Figure caption

Figure 1: Aspects of non-declarative memory. All the skills shown in white circles involve implicit memory. Those in the dark grey circle involve statistical learning, with the subset in the light gray circle involving verbal learning. The skills bounded by the dotted line (all involving sequence-learning) are all postulated to be impaired by the Procedural Deficit Hypothesis of Ullman and Pierpont (2005) (although other impairments may also be implicated).

