International Journal of Innovation and Research in Educational Sciences
Volume 2, Issue 5, ISSN (Online): 2349–5219

The Role of Statistical Learning in Overcoming Language Learning Difficulty in Adults

Majid Mehrabi
Ph.D. Candidate, Senior Lecturer at the
Department of English Language, Tonekabon Branch,
Islamic Azad University, Tonekabon, Iran.
Email: m.mehrabi@toniau.ac.ir

Firooz Sadighi
Professor of English Language& Linguistics, Shiraz Branch,
Islamic Azad University, Shiraz, Iran.

Abstract **– Saffran *et al.* suggest "infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language"; they motivate "innately based statistical learning mechanisms rather than innate knowledge" operating on "statistical properties of the language input" in order to allow the child to induce linguistic knowledge.**

Statistical learning approaches to language emphasize the richness of human communication: it is the primary source of data from which the child identifies patterns in their native language. In this Article, we are going to elaborate on the notion of statistical learning and try to show why it is necessary to understand the characteristics of such notion which arises from the computational human mind. This short, but rather general survey of the function of statistical learning will hopefully manifest, at least in part, some of the language learning difficulties in adults and may provide ideas for the practicing teachers as to how to tackle the issue.

Keywords **– Conditional Statistics, Distributional Statistics, Infant Learning, Learning Mechanism, Statistical Learning.**

I. Introduction

Statistical learning approaches to language emphasize the richness of human communication. It is the primary source of data from which the child identifies the patterns in their native language. No other animate being has the capacity to utilize a means of eliciting such system of communication other than humans. It is a human species – specific endowment. Statistical learning refers to the process of identifying units in the input guided by the statistical structure of the environment such as words or categories. Human child discovers what features of the input predict other features, and develops the ability to group features that are likely to co-occur and thus discover patterns. As such, associative learning is clearly regarded as an important component of statistical learning. It is a domain - general ability where learners can discover statistical relations in many different types of input of which language material is only one. But there are a variety of potential statistical relations to which learners could attend. Saffran *et al*.'s (1996b) experiments on word segmentation in infancy provide a concrete example of statistical learning. We will, in the remainder of this reflective investigation, examine statistical learning in more detail and focus on three breakthrough/enlightening questions. First, to what statistical features of the environment are learners sensitive? Second, how is statistical learning constrained? And three, how do the characteristics of the learning organism affect the outcome of statistical learning?

II. Environment and Sensitivity to Statistical Features

Statistical learning is guided by the statistical structure of and information in the environment. From a descriptive perspective, we can group these statistics into two broad categories of conditional statistics (CS) and distributional statistics (DS).

Conditional Statistics

In the statistical learning literature, transitional statistics is the most familiar one which provides a preliminary introduction to conditional statistics. The transitional probability is in fact the relationship between two items namely ***X*** and ***Y*** which can be formalized as the number of times the sequence ***X-Y*** occurs, divided by the number of times ***X*** occurs. Both infants and adults can use transitional probabilities to group items that are highly likely to co-occur (Aslin *et al.* 1998). Infants can use this to group syllables and segment words from the stream of fluent speech (Saffran *et al.* 1996a). As a matter of experimental demonstrations, infants are sensitive to transitional probabilities from as early as 2 months of age.(Kirkham *et al.* 2002). It is interesting to point out that transitional probability(TP) is not confined to adjacency of elements and many of the relations infants and adults alike learn involve regularities between non-adjacent elements. This is especially true of languages. While the definite article marker *the* predicts that a noun shall follow, the noun can follow several words later as in: the *big brown dog*. Several experiments have demonstrated that both infant and adult learners can detect non-adjacent TP's (Newport & Aslin 2004, Creel *et al.* 2004).

Although transitional probabilities are evidently informative of a trend, there are many different kinds of conditional statistics available to learners beyond TP. One such statistics may be drawn from co-occurrence probability where there exists the likelihood that two or more events occur simultaneously. But there is a difference between the two. While transitional probabilities assess sequential relationships, co-occurrence statistics measure simultaneous relations. Again, both infants and adults are sensitive to co-occurrence statistics (Chun & Jiang 1999, Younger & Fearing 1998). Thus, transitional probabilities are but one example of the kinds of conditional statistics to which learners are sensitive. This suggests that statistical learning may be applied in a wide variety of different learning situations and in theory, applied to similar teaching situations as well. Also, take note of the fact that from the two, conditional probabilities are much more useful to learners than co-occurrence because they are a more sensitive measure of the strength of the relation between two (or more) items (see Schultz and Gopnik 2004).

Distributional Statistics

An equally informative group of statistics to which learners attend is distributional statistics. Distributional statistics reflect the relative frequency of an event. It is arithmetic of percentage. As such, it reflects information about the central tendency and variability of a group of events. Even very young infants are sensitive to these kinds of distributional statistics (Dougherty & Haith 2002, Maye *et al.* 2002).

Distributional statistics have long been suggested to be important for various aspects of language learning (e.g. Reber & Lewis 1977). Indeed, distributional statistics may play a role in one of the most striking linguistic developments in the first year of life: infants' adaption to the phonemic structure of their native language. At birth, infants distinguish between phonemic contrasts not found in their native language. At one year of age, infants are primarily sensitive to those sounds that are phonemic (i.e., indicate a difference in meaning) in their native language (Werker& Tees 1984).

Another aspect of DS is the ability to identify the most common feature or pattern in the input through being sensitive to distributional information. This allows the learners to learn a pattern that regularly occurs, but is occasionally violated (Saffron & Thiessen 2003). But another aspect of DS is information about variability which can be thought of as a measure of whether the distributional probabilities of a set of two or more events are equivalent or skewed. High variability is achieved where all of the events have roughly equal distributional probabilities. In a situation where one of the events has a markedly higher probability, there is lower variability. Adult learners can be exquisitely sensitive to the variability in their environment (Mueller *et al.* 1974). Infants are also sensitive to variability in their environment, although they may respond differently than adults (Hudson Kam & Newport 2005). Therefore variability plays a particularly important role in learning. For example, when learning to identify meaning in speech, listeners must learn that some changes in the acoustic signal indicate a difference in meaning (as in *big* vs. *pig*). It is only acoustic information that indicates a difference in meaning. Other variables, such as two speakers uttering, is ineffectual.

III. ARE 'DS' AND 'CS' RELATED TO THE SAME KIND OF LEARNING MECHANISM?

As already stated, CS describe the strength of relation between two or more items while DS describe the central tendency or variability of a distribution of items. While they both entail learning from the statistical structure of the environment, an important question to ask is whether they are tracked by the same learning mechanisms? As with all questions relating to mechanism, no single approach will be definitive. Here, we will take up the two approaches of formal and bahaviourial in order to try to trace the query.

A formal approach emphasizes identifying the computations that learners perform. There are similarities between DS and CS at a formal level in that both kinds of statistics require learners to track at least a rough approximation of the frequency of events in the environment. Indeed, conditional probabilities can be thought of as a special case of distributional probabilities. A conditional probability is simply a context-sensitive distributional probability. Distributional probabilities track the likelihood of some event, **Y**. Conditional probabilities track how likely **Y** is to occur in a particular context: after **X** (Christiansen *et al.* 1998, Vallabha *et al.* 2007). It is however not clear which formal statistics or computations best approximate the statistical regularities to which learners are sensitive. Various experiments have been conducted in many situations with transitional probabilities, mutual information and other formal statistical indices of relatedness, however, there is likely no single answer to the question of which units of representation are the primitive units of computation. Different types of stimuli will entail different primitives, and even within the same type of input, learners can use different units as a function of the structure of the input (Saffran *et al.* 2005. Also read Aslin *et al.* 1998, Xu & Tenenbaum 2007, Redington *et al.*1998, Swingley 2005, Newport *et al.*2004, Aslin & Newport 2004).

A complementary second approach is focusing on behaviourial data. If sensitivity to different kinds of statistical information arises from different learning mechanisms, then there should be a divergence in the age at which sensitivity emerges, or sensitivity to one kind of input is shown but not to another. Of course, adults are sensitive to both CS and DS (Saffran *et al.* 1996b). By 8 month, infants are also sensitive to both CS and DS (Maye*et al*. 2002, Saffran *et al.* 1996a).

In summary, statistical learning refers to learning that is guided by the statistical structure of the environment. But as we have seen, there are a variety of potential statistical relations to which learners could attend. Even beyond the two broad types of statistical information – CS & DS – there are a multitude of potential relations available based on the elements of computation : for example, phonemes, syllables, words and phrases. How can learners possibly sort through this multitude of potential statistics, and discover useful relations? This is what we will address next.

IV. CONSTRAINTS ON STATISTICAL LEARNING

Pinker (1977) coins the term 'combinational explosion' to refer to the uneasy realm of statistical learning. One argument is that while there are, in principle, an infinite number of statistical relations a learner might attempt to track in the input, there are only a finite number of cases a learner experiences to determine which statistics are fruitful. What is of importance to take note of is that for learning to succeed, statistical learning must be constrained in a manner that not all statistics are equally likely to be considered.

A second argument for constraints on statistical learning relates to the world linguistic systems that despite surface dissimilarities share deep commonalities in the way they are organized (Pinker 1994). This is in compatibility with the central hypothesis of the UG tradition. The key prediction of Universal Grammar is that language learning is constrained in ways that are unique to languages. To be specific, it means that infants learn about language using innate knowledge or mechanisms that are domain-specific; cross-linguistic similarities are a result of these domain-specific constraints on language acquisition.

Domain-Specific vs. Domain-General Constraints on Statistical Learning

An alternative perspective posits that language is partly learned through domain - general statistical learning mechanisms. However, these mechanisms are constrained, such that not all relations are learned equally well (Fiser & Aslin 2005, Newport & Aslin 2000, Saffran 2003, Saffran & Thiessen 2003). What is important is that these constraints are not specific to language. As statistical learning is a domain-general process, operating on many different kinds of input, the constraints on statistical learning are also domain-general. On the basis of this framework, the cross-linguistic similarities are one source of evidence that can identify the constraints on statistical learning. The rationale for the identification of these similarities is that learners are not insensitive blank targets for the input. They prefer certain kinds of statistical relations shaped by generations of language learners. It is indeed a question of language survival and the “survival of the fittest” that adopts certain linguistic structures that ‘fit’ with the constraints on statistical learning and discards those which are not useful.

The sticking question, however, is that, is there evidence to suggest that statistical learning is constrained? The answer to this question is affirmative. Research with infants strongly supports constraints on statistical learning. Infants learn some patterns more easily than others (e.g. Saffran 2002, Saffran &Thiessen 2003). Research with adults, and computational simulations, suggest similar conclusions (e.g. Endress *et al.* 2005, Peperkamp *et al.* 2006).

Constraints and Simplification of the learning environment

Another claim in the statistical learning framework is that constraints on learning simplify the learning problem or alleviate the intensity of combinatorial explosion. An example of this kind of constraint is the embeddedness constraint proposed by Fisher and Aslin (2005).Using visual stimuli, they found that participants who had discovered a super ordinate structure were insensitive to the statistical relation between subordinate elements of the super structure. This shows that when learners attend to and discover a greater rule, they show insensitiveness towards the possible smaller regularities or relations and thus learning is done much easier because there is less or no need for minor computations to be carried out. This embeddedness constraint may be highly adaptive meaning it limits the number of potential computations a learner may perform.

Influence of Learner Characteristics on Statistical Learning

Identical input to identical learning mechanisms can lead to different outcomes as a function of the characteristics of the learner. Simple learning in other animals other than the human species is deliberately left out from our deliberations for its very limited scope. A great deal of evidenceis support to this notion. In the next part, we will examine how the characteristics of human learners influence statistical learning, with a particular focus on information processing, perception and prior experience.

Statistical Learning and the Influence of Information Processing, Perception and Prior experience

Statistical learning is considered to be a form of implicit learning, because learners frequently seem unaware of what, if anything, they have learned (Saffran *et al.* 1997, Stadler 1992). But, even implicit learning can be affected by information processing abilities such as (i) attentional control and (ii) working memory (e.g. Stadler1995, Baker *et al*.2004).

(i) Infants identify statistical relations more readily in stimuli that catch their attention (Thiessen*et al.* 2005). Learners appear to be greatly impaired when they are forced to divide their attention between two sources of input in the same modality, such as speech and tone (Toro *et al*. 2005).

(ii) Working memory too plays an important role in determining the statistics which learners are able to detect (Newport, 1998).

Perception and the modality of the input by the learner too have a significant role in learning. For example, when exposed to audio stimuli, listeners are quite adept at identifying sequential regularities: A occurs, then B, then C (Saffran*et al*. 1996a). Visual stimuli, however, exacts less adeptness. Yet, learners exhibit optimal adaption in tracking relations when items co-occur together (Conway & Christiansen 2005, Saffron 2002). Therefore, manner of perception is key to identification of patterns and learning. The relation between perception and statistical learning is bidirectional in that perception and statistical learning have reciprocal effect on one another. The amount of flexibility in allowing input to shape their subsequent perception is naturally more lax in children than adults due to greater previous entrenched experiences of adults.

A third characteristic of the learner that affects statistical learning is prior experience. What a learner knows affects subsequent learning. Repeated reference to the varied formations of item/s of information in the input pool acts positively in learning. For example, infants are sensitive to the cumulative statistical information in making word-object pairings(Yu & Smith 2007). Children also resort to several biases or adaptive assumptions to resolve or simplify the word-learning problem (Markman 1991). Some of these biases may be developed as a result of children’s sensitivity to statistical information in the environment. One such assumption is the shape bias which may have been developed by the assumption that words refer to categories of objects with the same shape. This bias seems to have developed as a function of children's experience (Landau *et al*. 1998). Essentially, through experience they detect that the words that they learn may refer to objects with similar shapes. Learning regularities like the shape bias, which constrain future hypotheses, occurs across several different domains as a function of the statistical regularities in the input (Kemp *et al.* 2007).

Previous experience constrains subsequent statistical learning (e.g. Curtin *et al.* 2005). These constrains are adaptive and are compatible with the characteristics of the input. In fact, statistical learning would be insufficient for many of the learning challenges a child faces if it were not shaped by previous experience. For example, transitional probabilities alone are not sufficient to identify word boundaries in fluent, natural speech. Learners also make use of phonotactic, rhythmic and other acoustic cues (e.g. Christiansen *et al.* 1998, Thiessen & Saffran 2003 , Yang 2004). Learners incorporate other cues such as stress which may signify the beginning of a word or otherwise in the stream of fluent speech to identify the function of these acoustic cues. This is a highly adaptive strategy but it has it pitfalls. Better adaption to one environment means lesser one to another (Best & McRoberts, 2003). This has implications for change in learning outcomes which comes about as a function of age.

Age-related changes as a constraint

As discussed above, Information processing abilities, perception and prior experience as human-specific constraints on statistical learning change with age. This explains one of the twists of language acquisition – Why is it that young infants are more successful in acquiring language than adults (Johnson & Newport 1989)? Here the idea of age (articulated as the critical age before puberty) has been emphasized. The supposition is that if a learner does not master language within this critical period, they are unlikely to ever achieve full linguistic competence even though they may be able to achieve native-like levels of fluency (Birdsong & Molis 2001). Those adults find it more difficult to acquire language than infants present an apparent paradox for theories of language acquisition that emphasize learning.

One argument to this paradox is to assert that statistical learning plays, at most, a peripheral role in language acquisition and as such is guided by mechanisms that are language-specific and available only to infants. Adults are unable to learn language as well as infants because they lack access to language-specific learning mechanisms (Chomsky 1995).

A second argument is to suggest that the constraints on statistical learning change with development as a function of the age and prior experience of the learner. One explanation is entrenchment hypothesis where the first language, once precipitated, can interfere with the process of the second language. Another explanation would be Newport's (1990) 'Less is More' hypothesis. According to this hypothesis, infants are better suited to learning language because of their information processing limitations on attention and memory. Newport attributes adult language learner's errors to frequent 'frozen forms'; that is utterances in which whole words or phrases are produced without appropriate awareness of their constituent words or morphemes. The superior information processing ability of the adult learner to perceive and remember allows them to store and process entire complex chunks of language such as phrases. In comparison young children may be able to process and store only component parts of linguistic stimuli. This is the advantage blessed to children to analyze language in appropriate component parts such as words against phrases, or morphemes rather than whole words. The less is more hypothesis illustrates a very important point. Infants and adults exposed to the same input may internalize very different representations over which to perform statistical computations, as a function of their prior experience, information-processing skills and perceptual abilities.

V. Conclusion

Statistical Learning is a sophisticated memory system that tracks frequency, distribution and co-occurrence. Despite the plethora of statistics available in the environment, learners are not overwhelmed by the wealth of information, especially infant learners. Astoundingly, infants show the capacity to integrate these different statistics in the process of developing/learning their native language. They resort to adaptive strategy to discover useful cues to language components such as stress as a useful means to realizing word boundaries. They also make use of other cues such as transitional probabilities in distinguishing word from non-word. When in conflict or combination, they identify the relationship or association through statistical computations. Infants have the capacity to detect the correlation between say, lexical stress and word onsets, through yet another cue termed as co-occurrence statistic. In the case of lexical stress, as an example, transitional probabilities help infants identify word boundaries, and co-occurrence statistic highlight where, in the newly discovered words, stress is occurring (Thiessen & Saffran, 2007).

Learners – whether infants or adults – must flexibly integrate varying kinds of statistical information throughout life; which they will. No single statistic will provide enough information to identify the structure of input as complex as language. Therefore, Statistical learning entails the process of identifying units in the input of this uniquely human endowment – Language.

Acknowledgment

We would like to appreciate the helpful suggestions made by Dr. Mohammad S. Bagheri at Shiraz I.A.U.

References

[1] Akhtar, N. (1999). Acquiring basic word order: Evidence for datadriven learning of syntactic structure. Journal of Child Language,26, 339–356.

[2] Akhtar, N., &Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. British Journal of Developmental Psychology, 14, 79–93.

[3] Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. Language and Speech, 46, 155–182.

[4] Aslin, R. N., &Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), Child Phonology: Vol. 2. Perception (pp. 67–96). New York: Academic Press.

[5] Aslin, R. N., Pisoni, D. B., Hennessy, B. L., &Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. Child Development, 52, 1135–1145.

[6] Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 9, 321–324.

[7] Austin, J. L. (1975). How to do Things with Words. Cambridge, MA: Harvard University Press.

[8] Baker *et al.*, (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15, 460-466.

[9] Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. Journal of Child Language, 32, 159–173.

[10] Barlow, J. A. (1997). A constraint-based account of syllable onsets: Evidence from developing systems. Unpublished doctoral dissertation, Indiana University, Bloomington.

[11] Bates, E., &MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), Mechanisms of Language Acquisition (pp. 157–194). Hillsdale, NJ: Lawrence Erlbaum Associates.

[12] Bialystok, E., &Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), Second Language Acquisition and the Critical Period Hypothesis (pp. 161–181). Mahwah, NJ: Lawrence Erlbaum Associates.

[13] Bailey, T., & Plunkett, K. (2002). Phonological specificity in early words. Cognitive Development, 17, 1267–1284.

[14] Berko, J. (1958). The child's learning of English morphology. Word, 14, 150–177.

[15] Berman, R. (Ed.). (2004b). Language Development across Childhood and Adolescence.

[16] Berman, R. A., &Dromi, E. (1984). On marking time without aspect in child language. Papers and Reports on Child Language Development, 23, 21–32.

[17] Best, C.T., McRoberts, G. W. (2003) Infant perception of nonnative contrasts that adults assimilate in different ways. *Language and Speech.*

[18] Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language.*

[19] Chen, L. & Kent, R. (2005). Consonant–vowel co-occurrence patterns in Mandarin-learning infants. Journal of Child Language, 32, 507–534.

[20] Chomsky, N., (1995). *The Minimalist Program.* Cambridge, MA : MIT Press.

[21] Christiansen, *et al*. (1998). Learning to segment speech using multiple cues : A connectionist model. *Language and Cognitive Processes*, 13, 221-268.

[22] Chun, M.M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10, 360-365.

[23] Clark, E. V. (1973). Nonlinguistic strategies and the acquisition of word meanings. Cognition, 2, 161–182.(2003). First Language Acquisition. (2nd ed.). Cambridge: Cambridge University Press.

[24] Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology :Learning, Memory, and Cognition*, 31, 24-39.

[25] Cooper, R. P., &Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. Child Development, 61, 1584–1595.

[26] Creel, *et. al.*, (2004). Distant melodies : Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology : Learning, memory & Cognition*, 30, 1119-1130.

[27] Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language, 2, 133–142.

[28] Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. Language and Speech, 20, 1–10.

[29] Demuth, K., Machobane, M., &Moloi, F. (2000). Learning word order constraints under conditions of object ellipsis. Linguistics, 38, 1–24.

[30] Dougherty, T. M., &Haith, M. M. (2002). Infants' use of constraints to speed information processing and to anticipate events. Infancy, 3, 457–473.

[31] Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., &Plunkett, K. (1996). Rethinking Innateness. Boston: MIT Press.

[32] Endress *et al.*, (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology.*

[33] Fennell, C. T. (2004). Infant attention to phonetic detail in word forms: Knowledge and familiarity effects. Unpublished doctoral dissertation, The University of British Columbia, Vancouver.

[34] Fennell, C. T., &Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. Language and Speech, 46, 245–264.

[35] Ferguson, C.A., & Farwell, C. B. (1975).Words and sounds in early language acquisition. Language, 51, 419–439. Reprinted with Appendix in W.S.-Y.

[36] Fisher, J.R., & Aslin, R.N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology.*

[37] Graf Estes, K., Evans, J. L., Alibali, M. W., &Saffran, J. R. (2007a). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science, 18, 254–260.

[38] Hudson Kam, C.L., & Newport, E.L. (2005). Regularizing unpredictable variation : The roles of Adult and Child learners in language formation and change. *Language Learning and Development.*

[39] Johnson, E. K. (2005) English-learning infants' representations of word forms with iambic stress. Infancy, 7, 99–109.

[40] Johnson, E. K., &Jusczyk, P. W. (2001). Word segmentation by 8-montholds: When speech cues count more than statistics. Journal of Memory and Language, 44, 548–567.

[41] Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive Psychology, 21, 60–99.

[42] Kemp *et al.*, (2007). Learning overhypotheses with hierarchial Bayesian models: *Developmental Science.*

[43] Kirkham *et al.*, (2002). Visual statistical learning in infancy : Evudence for a domain general learning mechanism. Cognition.

[44] Landau, B., Smith, L. B., & Jones, S. S. (1988). The Development, 3, 299–321.

[45] Lidz, J., Gleitman, H., &Gleitman, L. (2003a). Understanding how input matters: Verb learning and the footprint of universal grammar. Linguistic Review (2002). 19(1/2), 1–223.

[46] Lust, B. (2006). Child language: Acquisition and Growth. Cambridge: CambridgeUniversity Press.

[47] Markman, E. M. (1991). The whole-object, taxanomic, and mutual exclusivity assumptions as initial constraints on word meaning. In S. Gelman & J. Byrnes (Eds.), *Perspectives on Language and Thought : Interrelations in Development* (pp. 72-106*)*. Cambridge : CUP.

[48] Maye, J., & Weiss, D. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. Proceedings of the 27th Annual Boston University Conference on Language Development (pp. 508–518). Somerville, MA: Cascadilla Press.

[49] Maye, J., Weiss, D., &Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science, 11, 122–134.

[50] Maye, J., Werker, J. F., &Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82, B101–B111.

[51] Muller *et al.*, (1974). Verbal discrimination learning as a function of encoding variability. *Bulletin of the Psychonomic Society*, 4, 41-43.

[52] Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference, Child Development, 69, 295–308.

[53] Newport, E. L., &Aslin, R. N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S. Howell, S. Fish & T. Keith-Lucas (Eds.), Proceedings of the 24th Boston University Conference on Language Development (pp. 1–21). Somerville,MA: Cascadilla Press. (2004). Learning at a distance 1: Statistical learning of non-adjacent dependencies. *Cognitive Psycholoy,* 48, 127-162.

[54] Newport, E.L., (1998). Constraints on learning and their role in language acquisition. *Language Sciences,* 10, 147-172. (1990). Maturational constraints on language learning. *Cognitive Science,* 14, 11-28.

[55] Piaget, J. (1952). The Origins of Intelligence in Children. (M. Cook, Trans.) New York: Norton.

[56] Pinker, S. (1984). Language Learnability and Language Development. Cambridge, MA: Harvard University Press.

[57] Pinker, S., &Jackendoff, R. (2005). The faculty of language: what's special about it. Cognition, 95, 201–236.

[58] Pinker, S. (1994). The Language Instinct. New York: HarperCollins. (1989). Learnability and Cognition: The Acquisition of Verb-Argument Structure. Cambridge, MA: Harvard University Press. (1977). *How the mind works.* New York : W. W. Norton.

[59] Pinker, S., & Bloom, P. (1990). Natural language and natural selection. Behavioral and Brain Sciences, 13, 707–784.

[60] Reber, A.S., & Lewis, S. (1997). Implicit learning : An analysis of the form and structure of a body of tacit knowledge. *Cognition,* 5, 333-361.

[61] Redington *et al.,* (1998). Distributional information : A powerful cue for acquiring syntactic categories. *Cognitive Science: A Multidisciplinary Journal,* 22, 245-469.

[62] R. Kager, J. Pater & W. Zonneveld (Eds.), Constraints in Phonological Acquisition (pp. 109–157). Cambridge: Cambridge University Press.

[63] S affran, J. R. (2002). Constraints on statistical language learning. Journal of Memory and Language, 47, 172–196. (2003). Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science, 12, 110–114.

[64] Saffran *et al.,* (1997). Incidental language learning : Listening (and learning) out of the corner of your ear. *Psychological Science,* 8, 101-105. (2005). Changing the tune. The structure of the input affects infants' use of absolute and relative pitch. *Developmental Science,* 8, 1-7.

[65] Saffran, J. R., &Griepontrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. Developmental Psychology, 37, 74–85.

[66] Saffran, J. R., &Thiessen, E. D. (2003). Pattern induction by infant language learners. Developmental Psychology, 39, 484–494.

[67] Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by8-month-old infants. Science, 274, 1926–1928.

[68] Saffran, J. R., Newport, E. L., &Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. Journal of Memory and Language, 35, 606–621.

[69] Stadler, M.A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition,* 18, 318-327. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology : Learning, Memory, and Cognition,* 21, 674-685.

[70] Schultz, L.E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology,* 40, 162-176.

[71] Swingly, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology,* 50, 86-132.

[72] Thiessen *et al.,* (2005). Infant-directed speech facilities word segmentation. *Infancy,* 7, 53-71.

[73] Thiessen, E.D., & Saffran, J.R. (2007). Learning to learn : Infants' acquisition of stress-based strategies for word segmentation . *Language Learning and Development,* 7, 73-100.

[74] Toro *et al.,* (2005). Speech segmentation by statistical learning depends on attention. *Cognition,* 67, B25-B34.

[75] Vallabha *et al.,* (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences,* 104, 13273-13278.

[76] Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception : Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 7, 49-63.

[77] Xu, F., & Tenenbaum, J.B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science,* 10, 288-297.

[78] Yang, C.D. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Science,* 8, 451-456.

[79] Younger, B.A., & Fearing, D.D. (1998). Detecting correlations among form attributes : An object-examining test with infants. *Infant Behaviour and Development,* 21, 289-297.

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

AUTHOR'S PROFILE

Majid Mehrabi
Ph.D. Candidate, Senior Lecturer at the Department of English Language, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran.
Email: m.mehrabi@toniau.ac.ir