SSLA, **24**, 261–268
DOI: 10.1017.S0272263102002103

INPUT-DRIVEN LANGUAGE LEARNING

Michael Harrington and Simon Dennis
University of Queensland

Input-driven models provide an explicit and readily testable account of language learning. Although we share Ellis's view that the statistical structure of the linguistic environment is a crucial and, until recently, relatively neglected variable in language learning, we also recognize that the approach makes three assumptions about cognition and language learning that are not universally shared. The three assumptions concern (a) the language learner as an intuitive statistician, (b) the constraints on what constitute relevant surface cues, and (c) the redescription problem faced by any system that seeks to derive abstract grammatical relations from the frequency of co-occurring surface forms and functions. These are significant assumptions that must be established if input-driven models are to gain wider acceptance. We comment on these issues and briefly describe a distributed, instance-based approach that retains the key features of the input-driven account advocated by Ellis but that also addresses shortcomings of the current approaches.

Frequency plays a role in language learning both as an attribute of individual experience (task frequency) and as an attribute of the linguistic environment (feature distribution frequency). Both are observable entities that lend themselves to formal models of learning and processing. The effect of task frequency in learning is readily evident in the well-established relationship between practice and skill development (Newell & Rosenbloom, 1981). Practice of a particular form or structure will typically lead to increased facility in using that knowledge, an outcome Schmidt (1992) called "knowledge-strengthening." The role played by the distribution of linguistic features is more contentious. In the approach sketched out by Ellis, the frequency with which specific linguistic forms are distributed in the input determines how fast and well these

Address correspondence to: Michael Harrington, Linguistics Program, EMSAH, University of Queensland, Brisbane 4702, Australia; e-mail: mwharr@mailbox.uq.edu.au.

forms are learned. Task frequency and distribution frequency can, of course, be related. All things being equal, forms that appear in the input more often will be practiced more often. However, Ellis argues for a wider role for distributional frequency in learning—that is, as the basis for predicting the development of new language knowledge. Frequency is assumed to thus play a quantitative and qualitative, or “knowledge-changing,” role in language learning (Schmidt). This is a theoretically richer but far more controversial claim.

The input-driven approach advocated by Ellis is represented in connectionist models (Elman, 1993; Redington & Chater, 1998) and most widely known in SLA via the Competition Model (Sasaki, 1994; MacWhinney, 1997). These models ascribe a primary, explanatory role to distribution statistics: The probabilities of occurrence of specific form-function mappings in the input predict learning outcomes. The most controversial aspect of the approach is the assumption that these form-function mappings are learned solely via associative learning mechanisms (Lachter & Bever, 1988). These are deemed sufficient for language learning in general and for the development of higher order syntactic knowledge in particular. The approach denies the need for an abstract level of syntactic representation, which places it at odds with the standard view that assumes abstract syntactic rules are essential for capturing structure-sensitive syntactic relations, such as long-distance dependencies (Fodor & Pylyshyn, 1988). It has been generally assumed that this autonomous syntax is innately specified in important respects.

Although contentious, the input-driven approach advocated by Ellis is attractive in a number of ways. The associative learning processes assumed to underlie the approach provide a well-understood transition mechanism accounting for how the learner develops increasing knowledge of the language (Gregg, 1996). The approach also makes a serious commitment to understanding the structure of the learning environment and the role it plays in learning. The incorporation of the structure of the learning environment in the learning account distinguishes input-driven approaches from alternative accounts in which learner-internal linguistic and cognitive processes have been the primary focus and the nature of the input is assumed rather than tested (Smith, 1997). This is the case for generative approaches (Epstein, Flynn, & Martohardjono, 1996), where input has been characterized as a largely unspecified trigger of development, and for interactionist accounts that emphasize the learner’s modification of the input as the basis of learning (Long, 1996). Finally, the input-driven approach makes minimal assumptions about the role of innate capacities in the explanation of language learning (Elman et al., 1996).

In short, the input-driven approach provides an explicit and readily testable account of language learning. Although we share Ellis’s view that the statistical structure of the linguistic environment is a crucial and, until recently, relatively neglected variable in language learning, we also recognize that the approach makes three assumptions about cognition and language learning that are not universally shared. The three assumptions concern (a) the language learner as an intuitive statistician, (b) the constraints on what constitute relevant surface cues, and (c) the redescription problem faced by any system that seeks to derive abstract grammatical relations from the frequency of co-occurring surface forms and functions. These are fundamental assumptions that must be established if input-driven models of learning are to be successful.

THE LANGUAGE LEARNER AS AN INTUITIVE STATISTICIAN

Input-driven models of learning assume that the individual keeps track of the frequency of occurrence of specific distributional properties of the input. These properties include where a particular linguistic form occurs, the forms with which it co-occurs, and the frequency with which it occurs, both alone and in combination with other forms. The approach is based on a probabilistic model of human inferencing that assumes that decisions are made and actions undertaken based on a computation of probabilities and utilities. It emphasizes probability as a description and norm for human behavior (Gigerenzer & Murray, 1987). The individual learner is assumed to actively compute the probability of a given outcome based on available cue strengths and to make a judgment on interpretation that best satisfies the strengths of the various cues presented.

The evidence that individuals encode frequency as a matter of course comes from studies demonstrating that people can be extremely sensitive to frequency of occurrence information in performing tasks that make use of this information, either implicitly or explicitly (Hasher & Zacks, 1984). Research in the Competition Model framework has shown that the speed at which individuals learn and process a specific form-function relationship reflects the frequency of various cues in the language. For example, cues such as word order, case marking, and animacy can correlate to express agency (MacWhinney, 1998).[1]

Although there is considerable research indicating that individuals are sensitive to frequency of events, there is also evidence that use of frequency information is not the automatic occurrence that Ellis suggests. Tversky and Kahneman (1974) demonstrated that judgments of frequency can be affected by heuristics, or systematic biases in judgment. In the study, subjects were asked to judge whether each of five consonants appeared more frequently in the first or third position in English words. Nearly two-thirds of the subjects judged the first position to be more likely despite the fact that all five consonants appeared more often in the third position. The authors attributed the outcome to an availability heuristic that systematically biased the responses (e.g., retrieval of words that begin with *k*- is easier than retrieval of words in which -*k*- is the third letter).

Evidence for a role in judgment behavior for both frequency information and heuristics has yet to be adequately resolved. A recent proposal suggests an evolutionary basis for which type of knowledge might be employed (Hertwig, Hoffrage, & Martignon, 1999). Individuals may count features when it is necessary to meet adaptive goals, which is attested by evidence that infants are highly attuned to making frequency-based phonetic discriminations (Saffran, Aslin, & Newport, 1996) but not in settings that involve more strategic processing. Whatever merit this proposal may ultimately have, it is evident that individuals use both heuristics and frequency information in processing, and any account of language learning will have to accommodate this fact (Bever, 1992).

DEFINING THE PERCEPTUAL EVENT: WHAT CONSTRAINS CUE COUNTING?

The idea that the individual counts features in the environment, whether implicitly or explicitly, assumes that the relevant features are countable. As a model of learning alone, it is unconstrained. The world can be divided into an infinite number of discrete events, and there must be some way to focus on what is important. The physical and social environment certainly plays a role in framing what is salient and needs attending to. Brase, Cosmides, and Tooby (1998), for example, made an argument for natural units of analysis (e.g., faces and whole objects) that have conferred some adaptive advantage on humans. However, stipulating a natural unit of analysis for language is difficult precisely for the reasons that have traditionally been advanced for abstract syntax: What you see is not all that you get. The problem of cue constraint is directly related to the redescription problem to be discussed.

Cue-counting approaches must also deal with the role of attention in processing. Although the process appears to be largely implicit, it is also recognized that some attention must be paid to the stimuli for it to register, and this attention is more than just passing awareness. You are not able to provide an estimate of the number of red cars that passed you on the way to work this morning, but you were aware of them enough not to let any of them hit you. Research that has attempted to identify minimum levels of attention suggests that the individual must attend to a specific object for more than two seconds for it to register in such settings (Johnson, Peterson, Yap, & Rose, 1989). Input-driven approaches need to incorporate an attentional component in cue learning and specify how it interacts with frequency information.

LEARNING STRUCTURE: THE REDESCRIPTION PROBLEM

The greatest challenge facing input-driven models of learning is accounting for the development of abstract syntax based on the associative learning of the distribution of surface features alone. Since the early critiques of behaviorism by Chomsky (1959) and others, linguistics has been dismissive of any substantial role for associative learning in language development, a perspective readily evident in the SLA literature (see, e.g., Eubank & Gregg, 1995). To make a compelling case for associative learning as a knowledge-changing mechanism, input-driven models must demonstrate the learners' capacity to learn from the statistical regularities in the input.

There are two ways in which these regularities can be extracted from the input. The first way is driven by the simple statistical distribution of forms. Collocations are an example of such statistical regularities. These regularities can be extracted directly from the input through the calculation of relative frequencies of occurrence of forms. The second way in which regularities can be extracted from the input is indirect. These patterns represent higher order structural relations and can only be extracted by some kind of systematic transformation of the input data. Regularities such as the long-distance dependencies governing pronominal reference cannot be learned by reliance on co-occurrence statistics alone. As is evident in the example sentences, there is nothing in the distribution statistics of the words that reflects the structural relationship between *Mary* and *she* in (1).

(1) a. *Mary watched television before she had her dinner.*
b. *Before Mary had her dinner she watched television.*
c. *Before she had her dinner Mary watched television.*
d. **She watched television before Mary had her dinner.*

The ability to learn these relations depends on a mechanism for recoding, or the "redescription," of the raw input into more abstract representations (Clark & Karmiloff-Smith, 1993). The ultimate success of the approach will depend on demonstrating how this recoding occurs. An important connectionist landmark in this endeavor was Elman's work modeling subject-verb agreement and cross-clausal dependencies using a simple recurrent network (Elman, 1990, 1993). The use of a simple recurrent network enabled him to simulate temporal order, which permitted the network to learn simpler forms before more complex ones as well as incrementally increase the memory space available in learning. By starting small, the system was able to first develop the primitives of noun and verb from the input and then use the knowledge to drive the redescription into higher order structures. Elman thus established the capacity of a connectionist model to capture, at least in principle, the recoding of the statistical input into higher level representations (see Marcus, 1998, for a critique).

However, connectionist models like the simple recurrent network still have a number of serious problems as models of language learning. They are difficult to scale to substantive portions of a language, both in terms of the size of the vocabularies they can accommodate and the number of grammatical structures they can capture. As a consequence, testing is confined to small networks applied to strictly limited domains. Also, rather than use naturally occurring language, special training sets must be assembled (Ellis & Schmidt, 1998). These training sets are meant to capture the critical statistics of the language but typically are constructed based on the experimenter's assumptions about language statistics rather than rigorous corpus analysis. This may make scaling an unachievable goal and the models untestable.

CAPTURING STRUCTURE FROM INPUT: TESTING INPUT-DRIVEN APPROACHES

The tenability of input-driven accounts of language learning is ultimately an empirical question. One solution proposed to address the issues raised here is the development of hybrid models that assume abstract representations but also incorporate distributed representations (Jurafsky, 1996; Smolensky, 1995). Although these models combine the strengths of both the symbolic and input-driven approaches, they can also push back the problem of learning in that syntactic structure is assumed as a given (e.g., Jurafsky).

Instance-based approaches to cognition (Hintzman, 1986; Hummel & Holyoak, 1997) are an attractive alternative to these approaches for several reasons. Instance-based learning is based on the incorporation of individual instances of each learning event, and memory retrieval processes play a central role in processing and learning. Instance-based models provide a partial solution to the cue constraint problem in that they assume a string (e.g., utterance) is encoded as a unit. They also retain the desirable features of connectionist learning models—in particular, minimal assumptions about preexisting syntactic knowledge—but do not encounter scaling difficulties that beset connectionist models like Elman's simple recurrent network. In contrast to these models, training occurs by appending instances to memory rather than by applying a learning algorithm like the gradient descent optimization procedure. This aspect of instance-based accounts is particularly important for models of language learning in which scaling is a critical issue.

However, current instance-based theories of sentence processing are symbolic in nature and generally not capable of extracting grammatical class information from the input stream alone. Typically, they operate on part-of-speech tagged text for which significant syntactic groupings such as noun phrases have been identified prior to the application of the model (Argamon-Engelson, Dagan, & Krymolowski, 1999), or they require a database of already parsed text fragments from which to generalize (Scha, Bod, & Sima'an, 1999).

Instance-based approaches have been applied to a number of language phenomena, including the learning of prosodic (Nakisa & Plunkett, 1998) and lexical knowledge (Goldinger, 1998). We are developing an instance-based model of learning that learns structure directly from the statistical distribution of forms in natural linguistic input. It retains the strengths of the input-driven approach described by Ellis but does not assume preexisting structural representation nor is subject to the scaling problems faced by connectionist models. The Syntagmatic Paradigmatic (SP) model characterizes sentence processing and learning as the interaction of three memory systems that operate on distributed, instance-based knowledge representations (Dennis & Harrington, 2001). These memory systems are lexical, syntactic, and relational, each playing a distinct but complementary role in interpretation.[2] The interaction of three dedicated memory traces with distributed, instance-based knowledge representations means that the model can capture the systematic nature of language use from natural input without the need for the type of global syntactic representations that define the generative accounts of learning (Harrington & Dennis, 2000), and it does so in a way that makes direct use of frequency information in the input.

NOTES

1. Ellis emphasizes the individual as feature counter, but counting in the Hasher and Zacks framework and that in the case of the Competition Model are different. In the former, feature frequencies are recorded often unconsciously and retrieved directly (Hasher & Zacks, 1984). In contrast, the Competition Model assumes that individuals keep accurate tallies of cue validities for a particular variable (e.g., who is the subject of the sentence) as it correlates with other variables (occurring before the main verb, being animate, receiving nominative case marking, etc.). Thus, the task for the individual in the Competition Model framework is to track both individual event frequencies and their co-occurrences. It also requires a means by which to identify the various cues.

2. A lexical trace is defined as the paradigmatic associate of a word across the corpus. A syntactic trace is the set of syntagmatic associations within a sentence. A relational trace is the set of paradigmatic associations within a sentence. The relational trace is an extraction of the predicate information contained in the sentence. The model uses the retrieved syntactic and relational working-memory representations as a set of constraints on sentence interpretation and employs a gradient descent procedure to resolve these constraints. A detailed description of the model is available at: http://www.uq.edu.au/humanfactors/people/s.dennis.

REFERENCES

Argamon-Engelson, S., Dagan, I., & Krymolowski, Y. (1999). A memory-based approach to learning shallow natural language patterns. *Journal of Experimental and Theoretical Artificial Intelligence*, *11*, 369–390.

Bever, T. G. (1992). The demons and the beast: Modular and nodular kinds of knowledge. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 213–252). Mahwah, NJ: Erlbaum.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, *127*, 3–21.

Chomsky, N. (1959). Review of the book *Verbal Behavior*. *Language*, *35*, 26–58.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language*, *8*, 487–519.

Dennis, S. J., & Harrington, M. (2001). The syntagmatic paradigmatic model: A distributed instance-based model of sentence processing. In H. Isahara & Q. Ma (Eds.), *Proceedings of the second Workshop on Natural Language Processing and Neural Networks* (pp. 38–45). Tokyo: NLPNN2001.

Ellis, N. C., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*, 307–336.

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

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

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

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language research: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*, 677–749.

Eubank, L., & Gregg, R. (1995). "Et in amygdala ego"?: UG, (S)LA, and neurobiology. *Studies in Second Language Acquisition*, *17*, 35–57.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Mahwah, NJ: Erlbaum.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. In

W. C. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 49–81). San Diego, CA: Academic Press.

Harrington, M., & Dennis, S. J. (2000). A memory-based model of L2 sentence processing. Paper presented at the annual meeting of the American Association of Applied Linguistics, Vancouver, BC, Canada.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372–1388.

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, R. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 209–234). Oxford: Oxford University Press.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Johnson, M. K., Peterson, M. A., Yap, E. C., & Rose, P. M. (1989). Frequency judgments: The problem of defining the perceptual event. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 126–136.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition*, *28*, 195–208.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In A. M. de Groot & J. Kroll (Eds.), *Tutorials in bilingualism* (pp. 113–142). Mahwah, NJ: Erlbaum.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, *49*, 199–227.

Marcus, G. F. (1998). Can connectionism save constructivism? *Cognition*, *66*, 153–182.

Nakisa, R. C., & Plunkett, K. (1998). Evolution of a rapidly learned representation for speech. *Language and Cognitive Processes*, *13*, 105–127.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Mahwah, NJ: Erlbaum.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes*, *13*, 129–191.

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

Sasaki, Y. (1994). Paths of processing strategy transfers in learning Japanese and English as foreign languages: A Competition Model approach. *Studies in Second Language Acquisition*, *16*, 43–72.

Scha, R., Bod, R., & Sima'an, K. (1999). A memory-based model of syntactic analysis: Data oriented parsing. *Journal of Experimental and Theoretical Artificial Intelligence*, *11*, 409–440.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–385.

Smith, N. (1997). Structural eccentricities. *Glot International*, *2*(8), 7.

Smolensky, P. (1995). Reply: Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture. In C. MacDonald & G. MacDonald (Eds.), *Connectionism: Debates on psychological explanation*, *Vol. 2* (pp. 223–290). Oxford: Blackwell.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.