THE JOURNAL OF THE LEARNING SCIENCES, *14*(2), 301–309

How Children Learn the Meaning of Words and How LSA Does It (Too)

Paul Bloom. *How Children Learn the Meanings of Words,* Cambridge, MA: MIT Press, 2000, 312 pp., ISBN No. 0–262–02469–1 (hardcover).

Commentary by Max M. Louwerse and Matthew Ventura
Department of Psychology/Institute for Intelligent Systems
University of Memphis

For long psycholinguistics has tried to answer the question "how children learn the meaning of words?" Paul Bloom answers the question in his book with the same title. He argues that the mind does not have a module for language acquisition. Instead, children learn the meaning of words by a set of general cognitive abilities, including the ability to infer intentions, to perceive the world in terms of objects and events, and the ability to understand syntactic structures.

As discourse psychologists we have to emphasize that children likely do not learn the meaning of words through learning words. Instead, it seems more likely that the meaning of words is acquired through discourse. Discourse can take the form of conversations by oneself, with others or by others, situated in time and space. In fact, one could argue that children learn—at least in part—words solely by understanding their relation in context. This is how some computational techniques approximate the meaning of words, sentences, paragraphs and texts. One such common technique is Latent Semantic Analysis (LSA). LSA is a statistical, corpus-based, technique for representing world knowledge. It computes similarity comparisons for terms and documents by taking advantage of the fact that particular words occur in particular documents. LSA takes quantitative information about co-occurrences of words in documents (paragraphs and sentences) and translates this into an multidimensional space. In short, the input of LSA is a large co-

Correspondence and requests for reprints should be sent to Max M. Louwerse, Department of Psychology/Institute for Intelligent Systems, University of Memphis, 202 Psychology Bldg., Memphis, TN 38152. E-mail: mlouwers@memphis.edu

occurrence matrix that specifies the frequency of each word in a document. LSA maps each document and word into a lower dimensional space by using singular value decomposition (SVD). This way, the initially extremely large co-occurrence matrix is typically reduced to about 300 dimensions. Each word now becomes a weighted vector on the various dimensions. The semantic relationship between words can be estimated by taking the dot product (cosine) between two vectors. What is so special about LSA is that the semantic relatedness is not (only) determined by the relation between words but also by the words that accompany a word (see Landauer & Dumais, 1997). However, terms like *dog* and *animal* will have a high cosine value (are semantically highly related) not because they occur in the same documents together but because words that co-occur with one equally often co-occur with the other (see Landauer & Dumais, 1997; Landauer, Foltz & Laham, 1998). The method of statistically representing knowledge has proven to be useful in a range of studies. LSA has been used as an automated essay grader, comparing student essays with ideal essays (Landauer, Foltz & Laham, 1998). Similarly, it has been used in intelligent tutoring systems, comparing student answers with ideal answers in tutorials (Graesser et al., 2002; Hu et al., 2003). LSA can measure the coherence between successive sentences (Foltz, Kintsch, Landauer, 1998). It performs as well as students on the Test of English as a Foreign Language (or TOEFL) exams (Landauer & Dumais, 1997) and can even be used for understanding metaphors (Kintsch, 2000).

For the purposes of the discussion that follows let us assume that LSA is how all language processing works and that LSA can be considered a model of how children learn the meaning of words. Operating on that assumption we will explore five arguments advanced by Bloom in his book and show how they can be related to LSA. The five arguments involve a) anti-associationism, b) theory of mind, c) fast mapping, d) essentialism, and e) word spurt rejection.

ANTI-ASSOCIATIONISM

Bloom argues that a strict version of association cannot explain how children learn the meaning of words. This version of associationism goes back to empiricism and states that children learn the word *dog* by observing or thinking about dogs. The word *dog* and the corresponding thought become associated and thereby learned. One could take this view even further into operant conditioning, where associations between stimuli will get reinforced or punished. It is true that children's first words often refer to concrete objects that can be seen and touched. It is also true that words can be learned most easily by somebody pointing out what the word means. Bloom, like Quine (1960), argues however that the input for children is far from ideal, with the word not explicitly pointed out or with the object not being present. Even an associationist model that allows for noise is problematic, because one would expect the child to make incorrect mappings, but totally incorrect mappings are rarely found. According to Bloom, statistical covariation between word and percept is neither sufficient nor necessary for language acquisition, as learning the words for imaginary things suggests.

If we assume that cognition can be separated into a perceptual or external system and a reflective or internal system, Bloom discusses associationism in the light of the perceptual system, mapping words to things in the real world. Indeed, because we do not clutter our environment with 60,000 sticky-notes for all the words we know such a perceptual-system view would be ridiculous. But what if the child is able to learn associations in the reflective system, by bootstrapping meaning from co-occurrences of words? Some words might be learned by perceptual association, most others by reflective association. LSA does not have the senses to perceive the environment and cannot, therefore, make perceptual associations. However, the assumption behind LSA is that associations between words are built solely by their occurrence in the environment, which makes it a good model for the reflective system.

To illustrate this idea, imagine three levels. At the first level words share a direct co-occurrence that only takes place if the two words occur in the same event, like chair and table. Developmentally this relationship could be seen as the first association that gets constructed between concepts in the environment. The second level is the indirect relation that occurs when two words share a co-occurrence to the same word. For instance a word like *food* may directly co-occur with both *chair* and *table* outside events where *chair* and *table* mutual co-occur. Therefore words like *food* or words that fall in the overlap of co-occurrences of *chair* and *table* are examples of where *chair* and *table* become associated outside the co-occurrence of *table* and *chair.* The greater the number of third-party words between concepts, the greater the association. This relation could be seen as the second possible relationship that gets constructed between concepts in that the overlap between the neighborhood sets can be seen as an overlap of features that are classically proposed to be the foundations for associations between concepts. The third relation tries to capture additional information by measuring the distribution of near neighbor occurrences across multiple events or documents in a corpus. These neighborhood sets are supposed to be an additional meaning of the target word by their distribution of occurrences over time. This distribution of neighborhood occurrences in each event (i.e., document) can be seen as the semantic meaning of a word at time *t* by further exploiting the roles of features associated with each concept. For instance when comparing *bird* and *plane*, the feature *wings* is in both neighborhood sets, but there are also words not shared by both words (e.g., *engine, feathers*). However the latter words should give additional information about the meaning of the target word by measuring the co-occurrence of these words in each document. This relation should bring out differences between words in that each word has a unique neighborhood set with its unique distribution of occurrences over events.

Additionally this relation can also be seen as an implementation of a feature comparison model but now the occurrences of the features in events are represented as an additional meaning of the concept beyond feature overlap.

Similar clustering techniques have successfully been applied to infer syntactic categories. Various studies (Finch and Chater, 1992; Burgess & Lund, 1997) have shown that words belonging to different categories are clustered differently according to their preceding and following contexts. That is, the syntactic structure can be inferred by applying a high-dimensional method to clustering the co-occurrences of words in that corpus. These constraints can thus be imposed using frequency information alone. They allow for building specific associations thereby solving the problem classically put forward by Quine (1960) and revisited by Bloom of what to associate.

THEORY OF MIND

According to the theory of mind, children have the social cognition and pragmatic understanding to read the minds of others. For example, children at a very young age pick up that spiders are scary through parent's gestures and expressions of fear. Bloom reports on a recent study that shows that 12-month olds follow a robot's gaze only if the robot interacts with the child in a meaningful way (Johnson, Slaughter and Carey, 1998). This suggests that children interpret others as goal-directed agents, as intentional systems. Bloom makes a distinction between semantic and syntactic learning, but argues that they seem to be derived from a similar mechanism. That is, individual words in a sentence must be interpreted to convey any meaning of the sentence.

Although it is true that LSA cannot infer intentions, a co-occurrence technique can bootstrap semantics and syntax allowing for these abstractions. By syntactically building frames for certain verbs the child knows that "bananas can be eaten" and so can "sandwiches" and "cookies." By means of co-occurrences the child also knows that bananas, sandwiches and cookies are in some way related. As an example, the second nearest neighbor in the General Reading up to Third Grade (after *eating*) is *food* (cosine = .59; LSA space created from the TASA [Touchstone Applied Science Associates, Inc.] corpus used to develop *The Educator's Word Frequency Guide*; see http://lsa.colorado.edu). It is, therefore, not surprising that there is strong correlation between the child's syntactic development and its vocabulary development (Bloom, 2000: 53n). Furthermore, if we consider a symbolic representation of social cognitive and pragmatic information, LSA would in fact be able to account for this information and make it part of its semantic space. Thus theory of mind could be considered as just another cue that aids in the meaning of a word that comes from the frequency based associations.

FAST MAPPING

Not surprisingly, Bloom also discusses children's ability to understand the meaning of words without extensive exposure, without training and without feedback. Even after a one-month period both children and adults can retrieve the meaning of an object name that was presented to them incidentally, with no performance difference between children and adults. In fact, this fast mapping is a general cognitive ability that does not only occur in mapping words, but also facts. Bloom argues that this fast mapping is particularly salient in learning socially-transmitted information.

Obviously, there is little social transmission in LSA. So how does the child know what is important and what is not from an LSA point of view? First, the child relies on the frequency of words, phrases and structures. This exposure hypothesis can easily be tested by correlating the age of acquisition (Gilhooly & Logie, 1980) with the Thorndike-Lorge (1944) frequency of words ($r = -.375$, $p < .001$, $N = 4061$). This correlation shows that the content words a child acquires early are also the frequent content words in a language. To make this claim stronger, the top 5% of most frequent words are acquired before the child is 2.5 years old and the top 1% before the child is 1-year-old. The question then remains how the child would fast-map certain information it is not exposed to frequently? Typical for the fast mapping hypothesis is that children can even learn the meaning of a word by being exposed to that word a few times (Carey & Bartlett, 1978). It needs to be pointed out however that most fast mapping experiments test recognition rather than language production, something that makes the ability of fast mapping less remarkable. In addition, the fast mapping experiments all present multiple exposures to a novel word. But even if after only a few exposures a child can fast-map the meaning of a word, how can this be explained by LSA that fully relies on a intermediate frequency of a document (words with an extremely low frequency or an extreme high frequency are not very meaningful in LSA)? Our answer would first be that not every word the child learns is actually fast-mapped. Words like "philology" and "calculus" are acquired around the seventh year. It seems unlikely that even if a 12-month-old child is exposed to these words it will fast-map them. This means that there must be something special about the words that are fast mapped. It might be the case that those words have a special meaning at the time of exposure. For instance, imagine the child burning its hands at a stove. The meaning of "stove" will become repeated through additional cues, like the panic, mother's scorn, the redness of the stove top, and pain. In terms of LSA lots of documents relating to "stove" and "pain" will be added to the child's LSA space. Additionally, the other cues included in the stove-pain association serve as a familiar context that aids in the fast mapping process. Alternatively, if an unfamiliar word is associated with an unfamiliar context (e.g., a 5-year-old learning the meaning of "mitochondria"), the child will have trouble fast mapping because there is no familiarity to the words that are used to explain its meaning. Therefore, it can be argued that from an LSA perspective fast mapping does take place but only in contexts that include words that are frequently used by the child.

ESSENTIALISM

Essentialism concerns the relations between properties of an object and its classification. For instance, children show a preference to interpret new objects as falling into basic level categories. That is, children treat new words (e.g., *Fido*) as falling into intermediate categories of abstraction ("dog"), that are not too general ("animal") and not too specific ("bulldog"). Bloom argues that the proneness to simple categories is a bias not a conceptual limitation, in that one of the advantages of the basic category bias is that this strategy gives the optimal compromise between informativeness and distinctiveness.

The basic level category bias seems a challenge for LSA. However, it turns out that this part of essentialism can again easily be explained in terms of frequency. Informative and distinctive words just tend to be more frequent. The fact that more documents tend to contain "dog" (Thorndike-Lorge: 811) than "animal" (377) or "bulldog" (11) makes it more likely that dog is returned by LSA as a near neighbor. Not surprisingly, age of acquisition nicely follows this frequency trend (1.69 years; 2.22 years; 3.94 years). In addition to frequency of the word, the compromise between informativeness and distinctiveness that brings out the essence of a word can also be explained in terms of LSA's singular value decomposition. SVD reduces the co-occurrence matrix by filtering out very infrequent and very frequent relations, leaving the relations between the essential words.

A related bias is Bloom's *shape-as-cue-theory*. That is, children extend new nouns on the bases of shape rather than size, color, or texture. This is provocative because in many cases the size, color and texture are all characteristics that occur just as much with the noun as shape. For instance, when preschool children were told that an egg is "a dax," they generalized this over a football, rather than a nest or a loaf of bread (Baldwin, 1992). LSA would not be able to explain this in terms of frequency, with egg having a much higher frequency (Thorndike-Lorge: 847) than bread (423), football (203), and nest (113). A simple LSA comparison analysis that takes more into account than frequency alone does show remnants of the expected pattern. LSA comparisons between *egg, nest, football,* and *bread* show that a semantic relation between egg and football is present in the General Reading up to Third Grade but decreases from the 6th-grade to the 12th-Grade, suggesting that even the LSA space might be sensitive to the shape-as-cue-theory.

WORD SPURT REJECTION

The fifth and final argument from Bloom's book involves the word spurt, also called naming explosion, word burst or vocabulary spurt. Children learn words in a rate of less than a word a day. An average adult knows about 60,000 words. Some have argued that something magical must be happening in the vocabulary growth, because we acquire the 60,000 words 147 years earlier than should be expected. Furthermore, many have argued that this magic happens in the child's first months. Contrary to these theories, Bloom argues that a language spurt does not occur in children before 17-months-old. Secondly, Bloom argues that this phenomenon is far from magical. Better memory, better memory span and more experiences could explain the gradual increase in word learning between age 10 and 17. The reason that the word-learning rate slows down is because this is all that our immediate environment has to offer.

The absence of an early word spurt and the presence of a later one can be explained in LSA. Although predicting whether LSA learns semantic relations in a similar fashion as children do requires extensive computational simulations (Li, 2003), conceptually the process is straightforward. Given that the beginning state is low (no language capabilities) and the ending state is high (full language capabilities), the progress can either be described in a linear or nonlinear fashion (see Figure 1). Because in the most simplistic form LSA would learn no relation if one word is present, one if two words are present, three if three words are present, and six if four words are present; the curve that would emerge would have the shape of a sigmoid function, with the impression of a word spurt around the asymptote, similar to experimental findings.

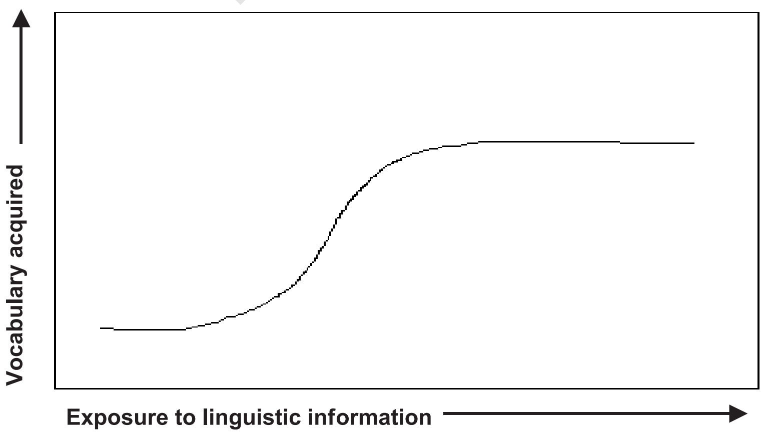

FIGURE 1 Word spurt in children and LSA.

No special abilities are needed when the LSA space is populated. In the beginning there are just not enough documents to allow for any semantic relations. Once there are enough documents, learning the meaning of words becomes child's play, while adding more documents that are similar in context will not generate new relations.

FOXES AND HEDGEHOGS

So do we really think that the mind of the child is just a large LSA space? We do not. Physical and psychological development, including memory and attention span, cannot be accounted for in LSA. Furthermore, because of the limitations of a symbolic representational system as the only representational system, LSA is at the most a model of part of the human mind. Having that said, we do believe that LSA explains certain key aspects of language processing. Furthermore, we believe that we can gain insight in some important psychological questions by using computational models like LSA.

Bloom ends his book with the distinction Berlin (1954) makes between hedgehogs and foxes to discriminate between different types of theories. Bloom's views are fox- rather than hedgehog-like. That is, instead of a hedgehog focusing on one aspect and explaining that one aspect in every single detail, Bloom's study is like a fox, doing many things and adapting itself according to the situation. Understanding word learning does not only come from studying each aspect independently (e.g., reflexes, low-level vision, phonological processing) but has to be seen in a broader context. In this lies an invitation to answer the question how children learn the meaning of words from perspectives other than psycholinguistics. We have gratefully accepted that invitation.

ACKNOWLEDGMENTS

This research was supported by NSF (IIS–0416128; REC–0106965) and IES (R3056020018–02). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF or IES.

REFERENCES

Baldwin, D. A. (1992). Clarifying the role of the shape assumption. *Journal of Experimental Child Psychology, 54*, 392–416.

Berlin, I. (1954). *The hedgehog and the fox*. London: Weidenfeld & Nicholson.

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT.

Bloom, P. (2001). Precis of “How children learn the meanings of words”. *Behavioral and Brain Sciences, 24*, 1095–1103.
Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 177–210.
Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15,* 17–29.
Finch, S., & Chater, N. (1992). Bootstrapping syntactic categories using unsupervised learning. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society* (pp. 820–825). Hillsdale, NJ. Lawrence Erlbaum Associates, Inc.
Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes, 25*, 285–307.
Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavior Research Methods and Instrumentation, 12*, 395–427.
Graesser, A. C., Hu, X., Olde, B. A., Ventura, M., Olney, A., Louwerse, M., Franceschetti, D. R. & Person, N. (2002) Implementing Latent Semantic Analysis in Learning Environments with Conversational Agents and Tutorial Dialog, *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, p. 37.
Hu, X., Cai, Z., Franceschetti, D., Penumatsa, P., Graessser, A. C., Louwerse, M. M., et al. (2003). LSA: The first dimension and dimensional weighting, *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.*
Johnson S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze following in 12 month-olds. *Developmental Science, 1*, 233–238.
Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin and Review.*, *7*, 257–266
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259–284.
Li, P. (2003). Language acquisition in a self-organizing neural network model. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp.115–149). Hove & Briton: Psychology Press.
Thorndike, E. L., & Lorge, I. (1944). *The Teacher’s Word Book of 30,000 Words.* New York: Teachers College, Columbia University.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.