Some Ideas about a Competence Theory of Vocabulary Acquisition and Its Possible Applications

Todd R. Davies
Psychology 211
Feb. 29, 1988

In this paper I would like to put forth some fairly sketchy ideas about how one might go about building a theory that would stand to vocabulary acquisition (and by extention to language acquisition generally) as grammars stand to parsing and syntax generation. In particular, I will be considering how an ideal agent might learn what the words of a language mean, i.e. a competence theory. Such a theory might help in the development of a psycholinguistic performance theory for how children learn words. In particular it might do this by identifying the constraints children may be attending to and, in ways that will be explained, by providing a possible standard for identifying errors or biases and for answering the most basic developmental questions about language acquisition.

In considering possible topics for this paper I originally had planned to consider an evaluation of the arguments for children constraining the meanings of new words, but found myself unable to disagree substantively with any of the points made in Markman (1987). Indeed, a brief consideration of arguments like those presented by Quine and Peirce really compells the view that constraints are brought to bear in language acquisition---it is hard for me to imagine anyone disagreeing with this. Likewise I find nothing to complain about in the experiments reported by Markman as support for the constraints of taxonomic organization and mutual exclusivity, so rather than considering these arguments in detail I thought it would be more interesting (and more fun) to ponder where one might go from here.

The problems about vocabulary acquisition that seem most salient to me after reading this paper are the following.

1. Even if we just consider nouns, it seems that other constraints are needed to explain how words are learned, besides the preferences for interpreting them as taxonomic categories and as mutually exclusive in meaning. If we imagine ourselves as children learning a new word, then even if we restrict ourselves to category interpretations for which we do not already have a term, we will be unable to eliminate, without further constraints, interpretations for "dog" that would essentially get its meaning wrong. Of course children apparently do not always get the interpretation right at first (they sometimes over- or under-generalize), but even apart from whether "dog" could refer to all medium-sized animals, or only to Rover, there is the question of why we should not interpret "dog" to mean "any object that is not Mommy," or "furry animal while the refrigerator door is open," or "brown object if inspected before tomorrow, or yellow object if inspected first thereafter." There are many constraints that are such common sense that it is hard to think of them, as well as some that we may be more aware of. So the first problem is what other constraints are necessary.

2. If we regard the goal of a competence theory for this task as the goal of specifying the constraints that should be imposed in acquiring vocabulary, and a procedure for applying them, then a natural question that arises is how to make the representation of these constraints and procedures more precise. One of the most important preconditions for the explosion of work on grammars was the invention of notations (especially production rule formalisms) in which to write them. Likewise, for vocabulary acquisition we might seek ways to encode the knowledge we assume to be present in a (possibly idealized) child that would allow it to learn words. Precision requires solving the problem of how knowledge like "Single nouns tend to refer to categories rather

than to thematic relations" could be grounded out in terms that would make it possible for a computer to apply it, and an attempt to solve this problem should reveal additional constraints and knowledge at each level of definition. It is an open question at just what level of detail the features used to define concepts like "category" could themselves be grounded out in a logical rule. One of the appeals of connectionism as a family of formalisms for representation is that it may be a better approach for relating low-level features like picture data to higher level concepts like "contiguous" and so forth. So the second problem is how to represent constraints.

3. A question that comes directly from Markman (1987) is how constraints, like taxonomic organization and mutual exclusivity, that may compete with each other are "traded off" in a way that makes one interpretation preferred. This is really the second order of business for a competence (or performance) theory of language acquisition after the identification of particular constraints. This is also a representational problem---one needs some kind of inferential calculus or procedure for deciding when an assumption must give way to other evidence or constraints. The third problem, then, is how constraints are applied and given priority.

4. Finally, it would be nice if we could say something about the extent to which constraints are known innately by the child as opposed to being learned by experience. Sometimes this question is not well posed because there are multiple possibilities for interaction between what we are biologically disposed to know and our experience, but in the case of language acquisition we may be able to say something about this issue in virtue of the fact that there are many different languages, and we initially learn just one of them. More will be said about this below. So the fourth problem is to what extent are constraints innate or learned.

I would like to discuss a strategy that may help us to solve the four problems outlined above. In so doing I am really assuming that it is possible to construct a competence theory that does not (to my knowledge) presently exist, and the feasibility of this proposal is an important question to keep in mind. Nonetheless, in the spirit of artificial intelligence and computational linguisitics, in which I have done some work, I will boldly suggest that what I propose can actually be done. What should be clear afterward is that if it is possible to do then the payoff should be quite grand.

The basic idea is as follows. The set of all natural languages, and each particular language, and each sample of language to which one child is exposed, all contain an objective source of information about the semantic constraints that actually apply to words. Many of these constraints are probabilistic, for instance, mass nouns tend to refer to superordinate categories (Markman, 1987), but do not always. A careful analysis of frequencies in, for instance, the data a child actually receives, should make it possible to build a list of statements about the relative likelihoods that particular constraints will be obeyed or disobeyed in a correct interpretation for a new word, given various pieces of additional evidence. A natural logic for building such a theory is probability theory, although the combinatorics involved in completely specifying knowledge in this form require making additional assumptions (usually involving conditional independence or minimal information) about how updating should occur, and these assumptions would have to be good ones for the data. That is, (and here the technical issues are really too involved to discuss in more detail) the set of probability statements should fairly characterize all those bits of knowledge that would be at substantial variance from the default assumptions (e.g. minimal information/maximum entropy) according to some error criterion.

The particular constraints children could attend to may apply at quite different levels of generality. For example, a target semantic rule might be of the form "Dogs are furry animals that bark," but on the way to inferring this the child almost surely requires more general rules or heuristics, of which the taxonomic organization and mutual exclusivity constraints are two examples. These constraints may have different likelihoods of being obeyed in each instance when a new word is uttered, and so one aspect of a competence theory would be to represent differences in frequency between constraints, possibly conditioned on various other pieces of evidence. But another goal of constraints is to narrow down the possible interpretations, so the informativeness of a constraint in this regard would be an important criterion for whether it is worth attending to. For instance, that words tend not to mean "Bob Hope" is quite true, but not terribly useful. Higher level constraints, or more general constraints, than the target semantic rule may often have the form of a functional dependency or determination rule (Ullman, 1982; Reference Note 1). For instance, a way to state the constraint that single nouns tend to refer to categories (as opposed to, say, thematic relations) would be the rule that the meanings of words, when the evidence is strong that they are single nouns, depend on the functional similarities among the instances to which they apply, or, equivalently, that function determines meaning. Probabilistically, as argued in Reference Note 2 and Davies (1988), this amounts to the assertion that the probability that two objects (broadly defined) are both instances of the same word, given that the objects share the same function, is high. Examples of other constraints are that nouns tend to refer to contiguous objects, nouns tend not to refer to an object at one time but not at another, nouns tend to refer to objects present when they are uttered (in the child's early experience), and nouns tend not to refer only to the negation of other nouns. If we are truly interested in building a theory that would tell someone who knows absolutely nothing about language (like a computer) enough to be able to learn words, then we really need all these very obvious constraints. Some of them, we can imagine, could have been otherwise. For instance, non-entrenched properties like Goodman's (1983) "grue" sound silly, but we could have a word like "caterfly" or "butterpillar", which would mean "caterpillar before a certain time, or butterfly after that time" and would make perfect sense. That we do not have such a word may reflect barriers to the formation of nouns that refer to things whose gross features change so markedly and so abruptly. For instance, a steer is not a "steer" when it has been sliced by the butcher; it is then "beef".

The general prediction is that the ease with which children learn word distinctions will be positively related to the prevalence in the child's language or data (or in languages generally) of instances of constraints that facilitate this learning. There is most likely a strong interactive relation between ease of learning and language structure. For instance, that children can learn part-whole relations more easily than class inclusion relations could be both a cause of and caused by the presence of mass nouns for superordinate categories. The mass nouns may be an aid to ease learning, or their presence could lead to confusion for count nouns, or both; the causal direction is often difficult to decipher, although in this case cross-linguistic data provide some clue (Markman, 1987). With an encoding of the constraints that objectively occur in language data, we could make predictions about when certain constraints would be favored over others by an ideal language learner, and we could compare childrens' performance to this standard as a way of detecting errors or biases in the child's application of constraints. Whether this can be done is a big question, but the task itself seems clear. We would like a set of constraints, and procedures for combination, such that acquisition of words given the samples of language encountered by the child is provably optimal, i.e. results in the fewest errors. Intuitively, we would expect that constraints in the language or data would be applied by children in a priority order that reflected the extent to which the constraints empirically held in the data themselves, and we would want to note cases in which this was violated.

If one could establish that some constraints children apply, and their procedures for applying them, were more consistent with a strategy that would make sense for languages as a whole than for the particular language and/or data the child was dealing with, then this would provide strong evidence that the constraints and procedure are innate in the child. This would require showing that the child applied constraints that did not make a lot of sense given just the data he/she had received, but would make sense if the child had some innate knowledge of languages in general. We would need to do a careful analysis of the objective occurrence of constraints in the data received by a child, as well as in languages generally, and try to determine whether there was enough information in the former so that the child could have learned the constraints through experience. Constraints that would be good for testing this would be ones that are obeyed by most languages, but are not obeyed either by that particular language being learned or by the sample of language to which the child has been exposed. I do not know enough about languages to propose such a test, but the method of frequency analysis discussed herein should allow us to find one.

To summarize, then, the probabilistic analysis of what constraints occur in languages and in data may allow us to solve the four problems described above for vocabulary acquisition theories. First, we would identify what constraints occur and are necessary to acquire words optimally. Second, we would encode this knowledge in a precise formalism (probability theory with functional dependencies). Third, we would make predictions for how these constraints would be applied based on the observed frequencies for various conditions. And fourth, we would compare predictions based on language data accessible to the child with those that would require knowledge of languages more broadly in trying to assess whether the knowledge children bring to bear in constraining word meanings is innate or learned.

REFERENCE NOTES

1. Davies, Todd R. (1985) Analogy. IN-CSLI-85-4, Center for the Study of Language and Information, Stanford University, Stanford, CA.

2. Russell, Stuart J. (1986) Analogical and Inductive Inference. Ph.D. Thesis, Dept. of Computer Science, Stanford University, Stanford, CA.

REFERENCES

Davies, Todd R. (1988) "Determination, Uniformity, and Relevance: Normative Criteria for Generalization and Reasoning by Analogy". To appear in Helman, David H. (Ed.) Analogical Reasoning. Dordrecht: D. Reidel.

Goodman, Nelson (1983) Fact, Fiction, and Forecast. Cambridge, MA: Harvard University Press.

Markman, Ellen M. (1987) "How Children Constrain Possible Word Meanings". In Neisser, Ulrich (Ed.) Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization. New York: Cambridge University Press, pp. 255-287.

Ullman, Jeffrey D. (1982) Principles of Database Systems. Rockville, MD: Computer Science Press.