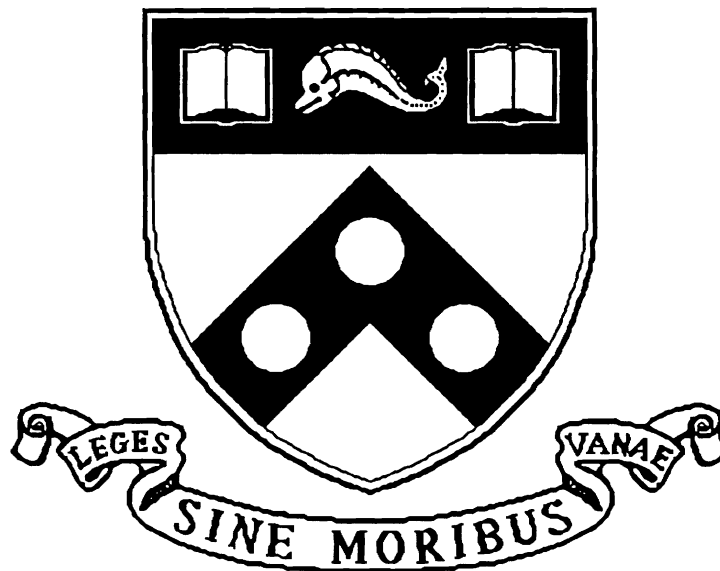


Acquisition Of Verb Categories

MS-CIS-92-76
LINC LAB 237

Mark Steedman



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

October 1992

ACQUISITION OF VERB CATEGORIES*

MARK STEEDMAN
University of Pennsylvania

October 15, 1992

ABSTRACT

The paper was delivered as a commentary upon Michael Brent's presentation "Acquisition of Subcategorization Frames Using Aggregated Evidence from Local Syntactic Cues" to the Penn Language Acquisition Conference, IRCS, University of Pennsylvania, 1992. It argues in support of using statistical techniques like Brent's to minimise the consequences of errors and misanalyses, but concludes that the case for believing that children acquire subcategorisations and other aspects of syntax on the basis of semantic and contextual cues remains strong.

The question of how children acquire lexical entries for verbs, and in particular their subcategorisation frames is one of the central questions concerning the child's acquisition of syntax. Its importance is enhanced by the recent tendency in theories of grammar to gravitate to a lexicalist position, and the role of verbs as the head of their clause. How do children do it, given the non-determinacy and automata-theoretic complexity of the syntax itself, and the unsystematic presentation and error-proneness of the linguistic data that

*Thanks to Lila Gleitman and Jeff Siskind for reading the draft. The research was supported in part by NSF grant nos. IRI90-18513, IRI90-16592, and CISE IIP, CDA 88-22719, DARPA grant no. N00014-90-J-1863, and ARO grant no. DAAL03-89-C0031.

they apparently have to make do with? Michael Brent’s paper in this volume shows how the statistical technique of binomial error estimation can be used to minimise the effect of contamination in the data available to the child language learner, arising either from errors in the input itself or errors in the child’s analyses of the input sentences. The technique is demonstrated by applying it to the sentences of a corpus of actual adult-child conversations, to derive subcategorisation frames for verbs from analyses based on imperfectly reliable local syntactic cues defined in terms of sequences of inflectional morphemes, function words and lexical NPs. As Brent points out, these two aspects of the work are quite independent: binomial error estimation could be used to minimise the influence of errors arising from imperfect analysis procedures of any kind at all, including those based on semantic and prosodic information, as well as syntactic. The present paper considers the part that all of these sources of information may play.

§1 SYNTAX

The specific application of this technique to low-level syntactic cues, rather than these richer sources of information, in this and the related work in Brent’s 1991 thesis can be argued to deliver two further important results. First, it demonstrates a practical technique that actually can be used to automatically build lexicons on the basis of large volumes of text. Although this point is not discussed in Brent’s present paper, it is worth emphasising. Hand-built dictionaries are inevitably very incomplete with respect to the exhaustive listing of subcategorisation properties that are needed for many computational applications. Techniques based on simplified syntactic properties which probabilistically “compile out” syntactic and semantic properties of what a linguist would regard as “the grammar”, and working on the basis of statistical properties of their distribution over a large corpus may well represent the *only* practicable possibility for automatically extending such dictionaries. Full-blown deterministic parsing of the corpora of the requisite size using linguistically respectable grammars and/or semantic interpretations and deterministic parsing, is impracticably expensive computationally, using existing techniques, to the extent that it is possible at all.

Second, the present study demonstrates the important fact that the information needed to determine verb subcategorisations actually is there in the

distribution of these very low-level properties in input of the kind that children are actually exposed to. For example, one of the apparent problems for acquisition of subcategorisation frames on the basis of syntactic information alone is the systematic ambiguity in all languages between subcategorised arguments and nonsubcategorised adjuncts, illustrated for English by the following pair of sentences:

- (1) a. We put Harry on the bus
b. We met Harry on the bus

How can the child avoid erroneously subcategorising *meet* like *put*? Brent points out that it doesn't matter if they do, because it is (presumably universally) the case that the relative frequency with which the sequence $V NP PP_{on}$ occurs will be significantly higher for verbs like *put* that subcategorise for NPs and *on* PPs than for those like *meet* which subcategorise for NP and only allow PP as an adjunct. Binomial error estimation is able to distinguish the two distributions, and reject the child's spurious evidence from analyses suggesting that *meet* subcategorises for the PP. Similar results seem to follow for spurious occurrences of subcategorisations arising from extraction, as in *who did you put on the bus*, which might appear otherwise to suggest that *put* might subcategorise for PP alone.

It is therefore reasonable to ask whether the child language learner actually makes use of such purely syntactic cues to learn the lexical categories of verbs. Here Brent is extremely cautious, and goes out of his way to acknowledge the possible involvement of prosodic and semantic cues as well. He notes in passing that there are a number of open questions that need answering before we can be quite comfortable with the assumption that the child is using the closed-class cues. The most important is that both the function words themselves and the cue sequences based on them are language-specific. The question arises of how the child can possibly catch on to the fact that *it*, *that* and *the* are cue words, much less that the sequence $V it the$ suggests that *V* is probably a ditransitive verb, while the sequence $V that the$ suggests that *V* is probably a complement verb. It is hard to see that there is any alternative to knowing, besides the set of possible subcategorisation frames, a) the precise syntactic significance of each closed class word as NP, Spec of CP, etc., and b) some statistics about possible corpora, including facts such

as that complement-taking verbs are more common than ditransitives.¹

I shall remain equally cautious in the face of such open questions, and certainly would not wish to claim that the child *cannot* be using such cues. However, as long as these questions remain open, it also remains unclear whether we have escaped what Brent identifies as the “chicken-or-egg problem” of apparently needing to know some syntax to apply this procedure. This suggests that there may be some point to asking ourselves what *other* resources the child could call upon, and in particular whether the two alternatives that Brent mentions, prosody and semantics, can help a child learn the first elements of syntax, including their first subcategorisations, in the face of the kind of uncertainties in the input which he identifies.

§2 PROSODY

Although Brent shows how misanalyses arising from the argument/adjunct ambiguity can be overcome, the consequences of some other quite similar sorts of ambiguity, such as that between prepositions and particles illustrated below, are not so easily eliminable by distribution-based methods, since verbs subcategorise for ambiguous items like *up* in both its guises:²

- (2) a. We rang up the girl
- b. We ran up the hill

In the case of particles and prepositions it seems intuitively highly likely that prosody disambiguates the two. Lederer and Kelly 1992 have shown that adults can reliably identify which of the two a speaker has uttered. They have shown similar effects for the argument/adjunct alternation. Kelly (1992 and this volume) presents results which suggest that a number of further apparent

¹It is not enough to assume that the child simply looks out for verbs followed by all possible sequences of cue words, classifying verbs as “*it+the* verbs”, “*that+the* verbs”, etc. Such a classification does not determine a subcategorisation.

²Brent suggests that spurious analyses of verbs like *ring* as subcategorising for PP can be eliminated by observing *sets* of subcategorisations, presumably meaning that we can reclassify verbs that have been assigned subcategorisations of both PP and NP+P. However, this is a distinct (and language-specific) complication to the proposal, and appears likely to conflict with the other uses that have been proposed for such sets. The prosodic cues discussed below would allow this particular complication to be eliminated from his account.

ambiguities are also correlated with prosodic distinctions. It is true that none of these prosodic discriminators are invariably present. Nor does it seem at all likely that *all* the relevant ambiguities are marked in this way. For example, I know of no evidence that the V-PP sequence arising from extraction in a, below, differs in any prosodic respect from that in b:

- (3) a. Who did you *put on the bus*
 b. Who did you *run up the hill* with

However, where the information is marked, it may well be reliable enough to be used as evidence under appropriate distribution-based techniques such as Brent's own, especially when we recall that adult's speech to children is characterised by exaggeration of normal intonation contours.

However, a word of caution is in order here. Adults speakers do not actually use intonation to indicate syntactic structure, but to convey the distinctions of discourse meaning that are variously described in terms of "focus", or of oppositions such as "topic/comment", "given/new" and the like. While the elements that are marked in this way correlate with syntactic structure, this is for semantic reasons, rather than for ease of processing. When adults exaggerate intonation contours in speaking to children, it is extremely unlikely that that they are using the intonational markers in any very different way. It is therefore quite possible that children are using this information as a semantic, rather than syntactic, cue, as part of the third strategy under consideration here.

§3 SEMANTICS

As soon as it was appreciated that even quite trivial classes of grammar cannot be learned by mere exposure to their stringsets, and that there appears to be little evidence that any more explicit guidance is provided by adults, it was obvious that some other source of information, "innate" in the sense that it is available to the child prelinguistically, must guide them in acquiring their grammar. As has often been pointed out, the only likely candidate is semantic interpretation or the related conceptual representation.³ However

³In the context of modern linguistics, the suggestion goes back at least to Chomsky 1965, pp.56-59 and Miller 1967. But of course it is a much older idea. See Pinker 1979 for

inadequate our formal (and even informal) grasp on the child's prelinguistic conceptualisation of the conversational situation, there can be no doubt that it has one, for even non-linguistic animals have that much. There can therefore be no doubt that this cognitive apparatus, for reasons which have nothing to do with language as such, partitions the world into functionally relevant "natural kinds" of the kind investigated by Landau in this volume, individual entities, including events, propositions, and such grammatically relevant notions as actual and potential participants and properties of those events, as well as the attitudes and attentional focus of other conversational participants. Since the main thing that syntax is *for* is for passing concepts around, the belief that syntactic structure keeps as close as possible to semantics, and that in both evolutionary and child language acquisition terms, the early development of syntax amounts to little more than hanging words onto the preexisting armatures of conceptual structure is so simple and probable as to amount to the null hypothesis.⁴

Of course, as Chomsky has repeatedly pointed out, this realisation gets us practically nowhere. We have such a poor grasp of the nature of the putative underlying conceptual structures that it is extremely difficult to even design experimental tests of it, (quite apart from the other difficulties that arise in doing experiments with prelinguistic children).⁵ Gleitman and others in the present volume have made considerable headway in the face of these difficulties, but there is a long way to go. For similar reasons to do with limitations on current knowledge, it does not seem to constrain syntactic theory in any very useful way. Right now, (and this is Chomsky's substantive point), the most reliable entry to the human system of language and symbolic cognition that we have comes from the linguists' phenomenological grasp of

a review of some proposed mechanisms, including the important computational work of Anderson 1977, and see Gleitman 1990 for some cogent warnings against the assumption that such semantic representations have their origin solely in present perception and the material world in any simple sense of that term.

⁴The use of the words "little more" rather than "nothing more" is important. It would not be surprising to find that some part of syntax – perhaps the observed constraints upon consistent orders across heads and complements – had its origin elsewhere than in semantics.

⁵I am not saying that logicians and my fellow computer scientists do not have interesting formalisms for representing conceptual structures. In fact these systems are the sole source of formal theoretical devices that linguists have to draw on. But as knowledge representation systems, none of them as yet seem particularly close to the human one.

the syntactic epiphenomenon, which has only just begun to look as though it is yielding some insight into the underlying conceptual structure.

Nevertheless, the claim that semantics is the precursor of syntax is not without content, and has consequences for the question at hand. In particular, it immediately entails that if we are asking ourselves why children do not classify *meet* as subcategorising for NP PP on the basis of sentences like 1b, *we met Harry on the bus*, then we are simply asking the wrong question. A child who learns this instance of this verb from this sentence must start from the knowledge that the denoted event is a meeting, and that this involves a transitive event concept. It simply never crosses the child's mind that *meet* might subcategorise like *put*, because the conceptual representation doesn't suggest that.

Once again, taking this position raises more questions than it answers. We are only just beginning to make sense of the complex mapping between surface grammatical roles like subject and object, and the underlying thematic roles that seem to be characteristic of the conceptual level (I am particularly thinking of recent work by Grimshaw 1990.) It also raises the question of whether the child's conceptual representation really can be used reliably in this fashion, which Pinker 1989 has called "semantic bootstrapping", and, if not, how the child can cope with its unreliability.

§4 SYNTACTIC AND SEMANTIC "BOOTSTRAPPING"

Gleitman 1990 argues very persuasively that the child must often find itself in a situation which is ambiguous with respect to the conceptual representation. To take one of her examples, a child who is being read a story from their picture-book about a fox and a rabbit may have insufficient information from their understanding of the story and from a picture of the fox running after the rabbit to tell whether an unknown verb in an adult sentence of the form "The fox is VERBing the rabbit" should be associated with the concept of chasing, or the concept of fleeing. In a number of elegant experiments, she and Landau and their colleagues have shown that children who are artificially placed in this situation identify whichever conceptual representation is consistent with the syntactic form of the sentence. Since this is the only information that appears to be available to the children as a basis for the decision, they argue that children are capable of using a process of

“syntactic bootstrapping” to aid them in learning the subcategorisations of verbs. Such a process requires the child to generalise from its existing partial knowledge of syntax and verbal subcategorisations, perhaps via the “linking rules” discussed by Pinker and his colleagues (cf. Gropen et al 1991).

Gleitman’s proposal is in principle entirely consistent with semantic bootstrapping in the sense outlined here. It is actually quite likely that the child’s conceptual representation isn’t much *more* underspecified than being ambiguous as to whether this situation is an instance of transitive chasing or transitive fleeing, and a few other equally relevant propositions. That is, we can probably assume that the child knows what it is to be “read a story” that the story is “about the fox trying to eat the rabbit”, and a lot of relevant facts such as that “to eat something you have to catch it”, “rabbits don’t like being eaten”, and so on. (If they *don’t* know stuff like this, then they may be in a position to learn the nouns, but probably not to learn the verbs.) In this case, there aren’t likely to be that many other possibilities, and syntactic bootstrapping may well reduce the set to one possible meaning. If so, it may well do so in the face of syntactic complications that are irrelevant to subcategorisations, such as the presence of modifiers or adjuncts. Their presence will not lead the child to include them in crazy subcategorisations that will stop it from acquiring the right meaning. It already *has* an appropriate meaning, as one of a number of alternatives that syntactic bootstrapping correctly disambiguates.

One piece of circumstantial evidence in support of this conjecture is that that adults work this way too. There is increasing experimental evidence that the adult sentence processing mechanism deals with the huge degree of nondeterminism that arises from natural grammars by appealing to meaning, filtering out the myriad spurious paths that the grammar permits on the basis of whether they make sense, both on the basis of sentence-internal semantics, and of reference and extension in the context. This semantic filtering of spurious paths which would otherwise overwhelm the computational resources of the processor has been claimed to go on continually at every point in the parsing process, with very fine “grain”, probably more or less word by word. (See Steedman & Altmann 1989 and Clifton & Ferreira 1989 for references and arguments pro and contra this proposal, which ultimately comes from computer science, particularly in work by Winograd 1972.)

Nevertheless, it may well be the case, as Gleitman suggests, that children are frequently much more at sea than this, and may even have much larger sets of propositions in mind, of which most or even all of which are irrelevant to the adult meaning. However, recent computational work by Siskind (1992, and this volume) shows that a process of intersecting such sets on successive encounters with the verb can be used to eliminate the spurious meanings.⁶

Of course, children are not adults, and neither are they mind readers, and a meaning that seems “appropriate” to them over a number of iterations of this process may not be the same as the adult’s. The child’s concept of “chasing” (we may imagine as an extension of Gleitman’s example) may be overspecifically restricted to an activity of attempting to catch by running. In this case, their own future use may be characterised by “undergeneralisation” – for example, they may be by unwillingness to agree that a similar scenario involving cars is chasing. There is of course a huge literature that has revealed the fine detail of this process.⁷ There are also instances of overgeneralisation, and possibly even more bizarre “complexes”, revealed in non-standard lexical meanings. There is also evidence that children predict new lexical entries that they have not actually encountered, via lexical rules such as the rule that generates causative verbs from certain adjectives, such as *cool*. This process may on occasion give rise to non-standard lexical causatives, as in *#It colds my hand*, either because of slightly non-standard lexical rules, or because standard rules are applied to slightly non-standard lexical entries. (See Bowerman 1982 and references therein.)

The way in which children successively modify non-standard lexical items to approximate the adult lexicon is the most challenging and least well-understood part of the process. But the undoubted fact that the processes of syntactic and semantic bootstrapping appear to iterate in this way suggests that together they may constitute the process by which children gain access

⁶Two problems which Siskind leaves open are the problem of polysemous verbs, and the problem that arises when the set of putative meanings derived from an occurrence of the verb are *all* spurious. Both of these eventualities will lead to empty intersections. One simple tactic that might serve to distinguish them and thereby be used to maintain a coherent lexicon would be to respond to an empty intersection by keeping *both* entries, relying on a tactic like binomial error estimation to distinguish between true polysemous lexical entries and spurious ones on distributional grounds.

⁷For example, Brown 1973, Bowerman 1973 and Clark 1973, and Carey 1982, the last including an extensive review.

to concepts which are not immediately available to pre-linguistic sensory-motor cognition, and may thereby provide the force behind the explosive change in cognitive abilities that coincides, both in evolutionary and in child-developmental terms, with the appearance of language.⁸ Computational models of the kind proposed by Brent and Siskind will continue to provide the only way in which theories of this process, such as syntactic, prosodic and semantic “bootstrapping”, can be developed and evaluated.

⁸See Vygotsky 1962 for for some early speculations on the nature of this process and its relation to Piagetian sensory-motor development, and see Oléron 1953 and Furth 1961 for some suggestive early studies of the effects of deprivation.

REFERENCES

- 1977 Anderson, John: 1977, 'Induction of Augmented Transition Networks', *Cognitive Science*, **1**, 125-157.
- 1973 Bowerman, Melissa: 1973, *Early Syntactic Development*, Cambridge University Press, Cambridge.
- 1982 Bowerman, Melissa: 1982, 'Reorganisational Processes in Lexical and Syntactic Development', in Eric Wanner and Lila Gleitman, (eds.), *Language Acquisition: the State of the Art*, Cambridge University Press, Cambridge, 319-346.
- 1991 Brent, Michael: 1991, *Automatic Acquisition of Subcategorisation Frames from Unrestricted English*, unpublished Ph.D dissertation, MIT, Cambridge MA.
- 1973 Brown, Roger: 1973, *A First Language: the Early Stages*, Harvard University Press, Cambridge MA.
- 1982 Carey, Susan: 1982, 'Semantic Development', in Eric Wanner and Lila Gleitman, (eds.), *Language Acquisition: the State of the Art*, Cambridge University Press, Cambridge, 347-389.
- 1973 Clark, Eve: 1973, 'What's in a word?' in T. Moore (ed.), *Cognitive Development and The Acquisition of Language*, Academic Press, New York.
- 1982 Clark, Eve: 1982, 'The Young Word Maker', in Eric Wanner and Lila Gleitman, (eds.), *Language Acquisition: the State of the Art*, Cambridge University Press, Cambridge, 390-428.
- 1965 Chomsky, Noam: 1965, *Aspects of the Theory of Syntax*, MIT Press, Cambridge MA.
- 1989 Clifton, Charles and Fernanda Ferreira: 1989, 'Ambiguity in Context', *Language and Cognitive Processes*, **4**, 77-104

- 1961 Furth, Hans: 1961, 'The Influence of Language on the Development of Concept Formation in Deaf Children', *Journal of Abnormal and Social Psychology*, **63**, 386-389.
- 1990 Gleitman, Lila: 1990, 'The Structural Source of Verb Meanings', *Language Acquisition*, **1**.
- 1990 Grimshaw, Jane: 1990, *Argument Structure*, MIT Press, Cambridge MA.
- 1991 Gropen, Jess, Steven Pinker, Micelle Hollander & Richard Goldberg: 1991, 'Affectedness and Direct Objects: The role of Lexical Semantics in the Acquisition of Verb Argument Structure', *Cognition*, **41**, 153-196.
- 1992 Kelly, Michael: 1992, 'Using Sound to Solve Syntactic Problems', *Psychological Review*, **99**, 349-364.
- 1992 Lederer, Anne & Michael Kelly: 1992, 'Prosodic Information for Syntactic Structure in Parental Speech', *Cognition* (incomplete reference).
- 1967 Miller, George: 1967, *The Psychology of Communication*, Penguin, Harmondsworth.
- 1953 Oléron, Paul: 1953, 'Conceptual Thinking of the Deaf', *American Annals of the Deaf*, **98**, 304-310.
- 1979 Pinker, Steven: 1979, 'Formal Models of Language Learning', *Cognition*, **7**, 217-283.
- 1989 Pinker, Steven: 1989, *Learnability and Cognition*, MIT Press, Cambridge.
- 1992 Siskind, Jeffrey: 1992, *Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition*, unpublished Ph.D dissertation, MIT Cambridge MA.
- 1989 Steedman, Mark and Gerry Altmann: 1989, 'Ambiguity in Context: a Reply', *Language and Cognitive Processes*, **4**, 105-122.
- 1962 Vygotsky, Lev: 1962, *Thought and Language*, MIT Press, Cambridge MA.

1972 Winograd, Terry: 1972, *Understanding Natural Language*, Academic Press, New York.