Language learning from fragmentary input

James R Hurford
Language Evolution and Computation Research Unit,
Department of Linguistics,
University of Edinburgh.
jim@ling.ed.ac.uk

Abstract

A model of vocabulary and grammar acquisition is presented. Two agents are involved in the simulation, a mother and a child. The mother is equipped from the outset with a substantial knowledge of language, in the form of two sets of rules. Her lexical rules map atomic meanings ('concepts') onto words. Her grammatical rules state generalizations about mappings between complex meanings and strings of words. The mother's rules allow her to utter strings of words expressing any meaning drawn at random from a large set. At the outset, the child has no such rules. The child does, however, share his mother's capacity for semantic representation; he has access to the same set of propositional representations, composed of the same atomic concepts. The mother utters wordstrings, which the child hears in full, but for each string, the child is made aware of only a small fragment of the mother's original meaning. From this exposure to wordstrings and small fragments of meaning, the child acquires a set of rules functionally equivalent to his mother's, and is capable of expressing the whole range of meanings with the same wordstrings as her. The child has fully acquired his mother's language, from data that is semantically highly degenerate. Early lexical acquisition is bootstrapped from observed correlations in the child's input. Grammar acquisition depends on some earlier acquired vocabulary. Later lexical acquisition takes advantage of acquired grammar rules. The whole process is informed by the learner trying to **make sense** of the data.

1 Introduction

There is an apparent paradox in learning to communicate. A creature that has acquired a communication code can retrieve meanings communicated by a signaller by using the acquired code to interpret the received signal. This is the great advantage of communication, that it allows one creature to know what is in the mind of another creature without the magic of telepathy. On the other hand, an immature creature in the process of acquiring a communication code needs to be given clear examples of meanings paired with signals in order to be able to learn the code.

Context helps mature hearers to retrieve meanings from signals. Hearers process utterances in a combined top-down and bottom-up manner. Top-down information comes from what is expected in the context of the speech situation (which may also include hypotheses retrieved from parts of the signal); bottom-up information comes from the signal. Top-down processing enables hearers to reconstruct noisy utterances, by filling in the details of the words or phonemes the speaker is likely to have uttered, from context-led expectations. Bottom-up information enables hearers to fill in details of the meanings a speaker intends to convey beyond what may be expected from the context of the speech situation, from knowledge of the communal code.

In normal speech situations, meanings are not entirely redundant (i.e. predictable from context). Children learn the code relating signals to meanings from observation of normal speech situations. But how can they do it, if all they observe is the redundant parts of meanings? The answer seems to lie in the fact that an element of meaning that is redundant in one speech situation is not necessarily redundant in another. The child can in principle begin to acquire a code by first acquiring a pairing between constantly observed parts of signals and reliably redundant elements of meaning. This partial knowledge of the code enables the child to engage in a progressively greater amount of top-down processing during subsequent observations of meaning-signal pairs. The feasibility of this strategy for learning a code is tested by computer simulations of a language learning situation. In this situation, there exists a prior code, used by adults, who communicate non-redundant messages to each other, and children observe this usage. The learning curve for language produced by these simulations resembles features of actual language acquisition, in particular the curve of vocabulary growth and the leap in competence associated with the so-called "syntax explosion".

2 The model and results

2.1 Semantic representations

Semantic representations are common to both mother and child. They are in a simple predicate logic format, without quantifiers, but with embedding of propositions as arguments of some predicates. Simple propositions contain either a 1-place, or a 2-place, or a 3-place predicate, each accompanied by the appropriate number of arguments. The arguments in simple propositions are constants, purporting to denote individual people and things. Examples are:

sing(fiona)
love(bertie, veronica)
give(alice, book, michael)

(I adopt the convention of giving semantic representations in lower case italics.) The arguments of all 1- and most 2-place predicates are always personal names, as are the 1st and 3rd arguments of 3-place predicates; the 2nd argument of a 3-place predicate always denotes an inanimate thing. Further, some 2-place predicates take a whole embedded proposition as 2nd argument, as in

believe(max, love(bertie, veronica))

For practical purposes, the degree of this recursive embedding of propositions inside each other was limited during the simulations, usually yielding a maximum depth of three propositions.

Such semantic representations are fully known to both mother and child. In particular, the child is able to distinguish a semantically well-formed proposition from any other assemblage of symbols. And, given an unstructured collection of atomic concepts (i.e. individual argument terms and predicate terms), the child can, if it is possible, construct a well-formed proposition from them. The child can also distinguish between the various argument slots; he can, for example, if required, build a proposition with a designated individual term in a designated argument slot (e.g. 'Agent' or 'Patient').

In the simulations reported here, an inventory of 1000 atomic concepts was used, distributed as follows: 400 concepts identifying persons, 400 concepts identifying inanimate objects, 90 1-place predicates, 90 2-place predicates, 10 3-place predicates, and 10 predicates of propositional attitude (e.g. *know*) (also 2-place). For convenience, these elements did not have English mnemonic labels. Given this large inventory of atomic concepts, an enormous set of different propositional meanings was made available.

At the beginning of simulations, only the mother has any means of public expression of these semantic representations.

2.2 The adult language

The language is stored in the heads of its adult speakers, in the form of grammars. Each individual is capable of storing the same set of rules. Simple lexical rules specify mappings between atomic semantic terms and individual words, as in

fiona ↔ `fiona`
know ↔ `know`
sing ↔ `sing`
book ↔ `book`

(By convention, words in the public language are given here in lower case typewriter font.) It was convenient in some simulations to use English words to represent both a semantic concept and its corresponding word, but no 'awareness' of this relationship was available to the simulated agents; the relationship between atomic meanings and words was entirely arbitrary.

Grammatical rules take the form shown in the following examples:

PRED(ARG1) ↔ `FORM1 FORM2`, where
  PRED ↔ `FORM2`
  ARG1 ↔ `FORM1`

PRED(ARG1, ARG2, ARG3) ↔
`FORM1 be FORM2 -en to FORM3 FORM4`, where
  PRED ↔ `FORM2`
  ARG1 ↔ `FORM1`
  ARG2 ↔ `FORM4`
  ARG3 ↔ `FORM3`

(By convention, upper case italics are used here for variables over semantic terms, and upper case typewriter font letters for variables over words of the public language.) Now, to explain the above example rules. The first clause of each rule (up to the 'where') specifies a type of proposition, in terms of the number of its arguments, and the double arrow expresses the mapping onto a type of word-string. The subsequent, indented, clauses express conditions on this mapping. For example, in the first rule, a condition is stated that whatever semantic term instantiates the semantic variable *PRED* must be mapped (by other rules in the grammar) onto whatever form instantiates the form variable `FORM2`. The conditions on each rule form an unordered set; all conditions on a rule must be satisfied for a proposition to be mapped onto a word-string. It will be noted that the right hand side of the second rule above contains certain non-variable items, `be, -en, to`. These are grammatical function words; in this case `be` and `-en` signify a passive construction, and `to` is a marker of indirect object.

The mother knows 14 such grammatical rules, representing analogues of English intransitive, transitive active, transitive passive, ditransitive active double object, ditransitive active dative, ditransitive passive double object, and ditransitive passive dative, each with a tensed and an untensed version. Examples of particular sentences generated by these rules are:

Meaning = *ramble(george)*

```
george ramble
george tense ramble
```

Meaning = *love(john, mary)*

```
john love mary
john tense love mary
mary be love -en by john
mary tense be love -en by john
```

Meaning = *give(max, book, zoe)*

```
max give zoe book
max tense give zoe book
max give book to zoe
max tense give book to zoe
zoe be give -en book by max
zoe tense be give -en book by max
book give -en to zoe by max
book tense give -en to zoe by max
```

Meaning = *know(sue, hate(max, joe))*

```
sue know max hate joe
sue know joe be hate -en by max
```

etc.

It will be noted that, at the sentential level, the meaning-form mapping is one-to-many. Although sentential paraphrases exist, there is no sentential ambiguity; and there are no lexical homonyms and no lexical synonyms. These are simplifying factors. The form in which rules are expressed here makes no mention of any non-semantic or non-phonetic (i.e. autonomous syntactic) categories. This is another simplification. In reality the form-meaning pairings that constitute a language system are constrained by non-semantic categorizations, expressible (in the simpler cases) in the familiar grammatical terminology of Noun, Verb, Adjective, Noun Phrase, Verb Phrase, and the like. In many other ways, the grammar rules given here are very simple and crude, by comparison with the rules of many linguistic theories or of many NLP systems, such as machine translation systems. The concern here is to explore how in principle grammatical rules can be acquired from highly incomplete information about meaning. As very few such studies have been done, it will be useful to start with such a simple grammatical schematism.

2.3 Mother speaks to child

In the simulations, at each cycle, a whole propositional meaning was chosen at random. The initial choice was between depths of embedding, with depths of 0, 1, and 2 being equiprobable. A complex proposition of depth n ($n > 0$) was always a 2-place predication, with one atomic argument (identifying a person) and one propositional, argument of depth $n - 1$. For 0-degree, i.e. simple, propositions, the choice was between 1-place, 2-place and 3-place predications, with each type being equiprobable. Once a proposition type had been chosen, the atomic meanings contained in them were randomly selected from a Zipfian distribution, in which frequency is inversely proportional to rank. Thus for example, given 400 personal names, the frequencies of the first and last-ranked were in the proportion 1 to 1/400; given 20 3-place predicates, the frequencies of the first and last-ranked were in the proportion 1 to 1/20.

Having selected a specific meaning, the mother selected, again at random, a way of expressing this proposition, according to her internalized grammar, outlined above. For example, given the proposition *give(max, book, zoe)*, the mother might, with equal probability, have uttered any one of the eight wordstrings given above.

The child heard the whole wordstring (unaffected by any noise), and was also allowed to 'observe' a single atomic element of meaning picked at random from the proposition. For example, if the mother were to say `max tense give book to zoe`, the child might be given the single concept *max*. The child was also given information about the role which the given concept plays in the original proposition, e.g. whether Predicate, Agent, Patient, or Beneficiary. Thus, as far as meaning is concerned, all the child received, with each utterance spoken to him, is a pair, such as for example ⟨ Agent, *max*⟩.

It is a matter of logic that if a child is to acquire knowledge of meaning-form mappings from experience, then at least some information about meaning must be available from the context of situation. From the viewpoint of the message being communicated between speakers, any element of meaning available from the context of situation is redundant information.

2.4 Early lexical acquisition

At first, having no grammar of any kind, the child could not retrieve any meaning from any utterance that was spoken to him. The simulated child was endowed with a facility for storing memories of sentences, paired with their associated fragments of meaning . Using the above example again, the child might store:

⟨Agent, *max*⟩ : `max tense give book to zoe`

On another occasion, he might hear `fiona love max`, and be given the semantic fragment ⟨Patient, *max*⟩, and would thus store:

⟨Patient, *max*⟩ : `fiona love max`.

Every so often, the child reviewed his store, searching for cases in which a criterial number of sentences involving the same concept all contained the same word. When he found such a case, the child constructed a lexical rule of the sort

max ↔ `max`

Having constructed a rule, the triggering stored memories of sentence-concept associations are deleted. For the experiments conducted here, a criterial number of 3 instances of sentences containing a common word, was enough to give very reliable results. Very occasionally, a run would end with a lexical mistake, (i.e. a lexical rule discrepant with his mother's grammar) as when the child internalized a lexical rule such as

george ↔ `mary`

This could happen if a significant number of sentences about George happened to contain the word `Mary`, but such occurrences were very rare. Occasionally children make such mistakes, perhaps getting people's names mixed up.

The child was credited with special 'innate' knowledge about unstressed grammatical function words (or functional morphemes), such as `tense`, passive-marking `be` and `-en`, and the `to` signalling indirect objects. It is to be emphasized that the child was not credited with innate knowledge of the meanings, functions or grammatical distribution of these words. It was simply assumed that in the mother's output utterances, such words would be unstressed, and the child did not count or heed such unstressed words when surveying his store of sentences associated with semantic fragments, for the purpose of identifying criterially frequent correspondences.

In the simulations, if new lexical items are only acquired by this early induction procedure, the growth of the child's vocabulary is slow. With a Zipfian distribution of atomic concepts, vocabulary growth by this procedure is close to linear; if, unrealistically, all atomic concepts are made equiprobable, the growth of vocabulary is faster in early stages, but decelerates later. The two curves are compared in Figure 1.

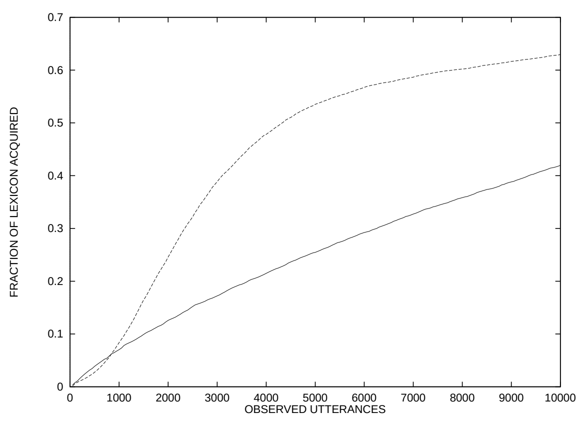

Figure 1. Vocabulary growth by early mechanism. Lower curve is with Zipfian distribution of atomic concepts. Upper curve is with non-Zipfian distribution of atomic concepts. Averaged over 10 runs.

2.5 Acquiring grammar rules

As soon as the child has acquired some lexical rules, by the 'guessing' procedure described above, he is able to retrieve more semantic information than before from sentences his mother speaks to him. Consider an example in which the child knows the three associations

max ↔ `max`
love ↔ `love`
fiona ↔ `fiona`

He is now given as input:
⟨ Agent, *fiona* ⟩ : `fiona love max`

The child gets one concept, namely *fiona*, both from the utterance and from the redundant context of situation, and two more concepts from the utterance itself, namely *max* and *love*; the child is also able to observe that *fiona* fills the Agent role. From this information the child is able to construct a single proposition involving all and only the presented concepts. The constructed proposition is *love(fiona, max)*.

Having unambiguously retrieved a whole proposition from an utterance, the simulated child makes a bold inductive leap from this instance to a generalization mapping all propositional meanings of the same type as the one reconstructed to corresponding sentences of the type from which it was retrieved. That is, from the token association

love(fiona, max) ↔ `fiona love max`

the child constructs a general grammatical rule by replacing the constants with appropriate variables, and listing the conditions on this rule in terms of generalizations from the particular lexical mappings that enabled retrieval of this proposition. The semantic variables apply to positions in the proposition, e.g. *PREDICATE*, *ARGUMENT-1*, and the linguistic variables apply to positions in the sentence, e.g. *FORM-1*, *FORM-2*. In this case, the constructed rule would be:

PRED(ARG1,ARG2) ↔ `FORM1 FORM2 FORM3`, where
PRED ↔ `FORM2`
ARG1 ↔ `FORM1`
ARG2 ↔ `FORM3`

This rule, acquired from exposure to strings and partial meanings, is identical to one of the mother's rules.

In constructing such a rule, the child treats recognized unstressed function words as constants in the wordstring, rather than as variables for which conditions on their mappings to concepts must be listed. So, given the same vocabulary as above, if the child had observed

⟨ Agent, *max* ⟩ : `fiona tense be love -en by` `max`

he would have constructed the rule

PRED(ARG1,ARG2) ↔ `FORM1 tense be FORM2 -en by FORM3`, where
  PRED ↔ `FORM2`
  ARG1 ↔ `FORM3`
  ARG2 ↔ `FORM1`

The "bold inductive leap" described above is obviously a simplification. Probably, as an account of real language acquisition by children, a statistical element needs to be added. That is, a child does not generalize from a single example, but from a criterially sufficient set of examples. I suggest, however, that such a mechanism of generalization, from particular (sets of) observed utterances paired with particular (sets of) reconstructed meanings, to rules stating generalizations over such pairings, is at the heart of grammar acquisition. The claim is that children's acquisition of language centrally involves **attempting to understand**, i.e. to retrieve meaning from, what is said to them, and generalization from success at such attempts. This is not to deny that other factors, such as more elaborate constraints on the form of grammars than are entertained here, also play some part.

The 2-place predication, expressed by a transitive sentence, just illustrated, shows the need for 'role' information to be available to the learner. If the learner knew that the intended message involved the three concepts *max*, *love* and *fiona*, but did not know which of *max* or *fiona* was the Agent, he would be able to construct two propositions, rather than just one, from the observed utterance. The child could then not safely generalize to a rule involving a single type of proposition. Conceivably, real children are equipped with some prior disposition to assume that words referring to Agents will tend, in 'unmarked' cases, to appear earlier in a sentence, but such a possibility is not explored here. However, another strategy for deriving clear role-assignments in reconstructed propositions is explored, as described below, after a note about reflexives.

A special difficulty exists with reflexive 2-place predications, such as *love(max, max)*, in which the Agent is identical to the Patient. If an adult were to express this proposition to a child with the sentence `max love max`, with the child also able to observe, say, ⟨ Patient, *max*⟩, tre child could only retrieve one (correct) proposition, but would nevertheless be able to infer two possible generalizations about the grammatical expression of 2-place predications. In one generalization, the Patient occupies the sentence-initial, preverbal position, and hence the Agent occupies the sentence-final, postverbal position; in the other generalization, these positions of Agent and Patient are switched. Both generalizations are valid on the basis of `max love max` meaning *love(max, max)*, but only one is correct for `mary love john` meaning *love(mary, john)*. In these simulations, this difficulty was overcome by the ad hoc device of not allowing the child to construct grammatical rules on the basis of examples containing repeated concepts. (Kirby, in two forthcoming papers, also excludes reflexive propositions from the data used by his grammar inducer for exactly the same reasons.) In real language acquisition, reflexive propositions do not present this problem, as real languages typically avoid using the same content word (e.g. a proper name) more than once in a sentence, preferring instead to use pronouns, as in "John loves himself".

In these simulations, the only semantic information given to the learner from the context of situation is a single atomic concept, together with information about its role (e.g. Predicate, Agent, Patient) in the intended message. As explained above, this latter information is crucial in inducing rules for expressing 2-place predications. But with 3-place predications, a similar problem arises for the learner, which cannot be solved if only a single concept is given from the context of situation. For example, assume that the child has learned the meanings of the words `max, fiona, book` and `give`, and now observes:

⟨ Agent, *max*⟩ : `max give fiona book`

The concepts *max*, *book*, *give* and *fiona* can be retrieved from the utterance by lexical lookup. Any proposition involving just these four concepts must be a 3-place predication, and it is known from the context of situation that the first (i.e. Agent) argument place is to be filled by *max*. But the assignment of *book* and *fiona* to the other (Patient and Beneficiary) argument slots is underdetermined by the evidence available from the utterance and its context of situation. Here, I invoke the prior knowledge of the child about the selectional restrictions on a 3-place predicate, such as *give*. It is assumed that the child knows that the Patient slot in any such predication is taken by an inanimate object, and that the Beneficiary slot is taken by a person. As the child also knows that *book* is inanimate, and that *fiona* is a person, the only reconstructible proposition from the given concepts is *give(max, book, fiona)*. Having constructed this proposition, the child makes the inductive leap, as before, to a general grammatical rule for the expression of all 3-place predications by a wordstring of the type used in the observed utterance.

In the cases so far discussed, a grammatical rule was induced on the basis of an utterance in which the learner knew the meanings of all the content words. It is clear, however, that if the learner knows the meanings of all the content words except one, and the meaning of this content word happens to be available from the context of situation (and does not repeat any of the meanings already detectable from the utterance), then a proposition can also be unambiguously reconstructed. Assume, for example, that the learner already knows the meaning of `fiona`, but does not know the meaning of `sing`. He may now be given

⟨ Predicate, *sing*⟩ : `fiona sing`

From this, knowing that `sing` is not a function word, and therefore a content word, to which some conceptual meaning can be assigned, the child can construct the proposition *sing(fiona)*, and then induce a general rule for the expression of 1-place predications. Here the child has simultaneously acquired a new piece of lexical knowledge (that `sing` means *sing*), and a general grammatical rule.

To put this situation in terms of a real learner's experience of real English, it would be like a situation in which the child knew the meaning of the name `fiona`, i.e. knew to whom that name was conventionally attached, and also had a concept of singing as an activity involving a single participant. Such a child could distinguish in her mental representations between a person singing, and, say, a person humming or jumping, but would not, as yet, have learned any word to express the concept of singing. Someone now says to the child "Fiona is singing", in a situation where the child can hear that there is some singing going on (say in the next room), but does not know who is singing. The child is, furthermore, made aware that the singing is what this utterance is about. The child puts together the knowledge that someone is singing, and that an utterance has just been made about this fact, with the person-concept *fiona* derived by lexical lookup from the utterance, and concludes that the utterance is conveying the message *sing(fiona)*.

Simulations were run with the grammar and lexical acquisition procedures so far described, with 1000 atomic concepts, distributed as described earlier, and with the mother using 14 different grammatical rules to express a variety of 1-place, 2-place and 3-place predications, sometimes with whole propositions embedded as objects of 2-place predicates. While the lexical acquisition proceeded at a slow steady pace, as shown earlier (Fig.1), the acquisition of grammar rules was, in terms of a proportion of the total number of facts to be acquired, relatively fast. It is not possible to make a straight comparison between the rate of acquisition of a vocabulary of 1000 items and the rate of acquisition of just 14 grammatical rules (like comparing the proverbial apples and oranges). But, if one plots the acquisition of grammatical rules on the same graph as the acquisition of word-meanings, in terms of a proportion of what is to be learned (1000 word-meanings or 14 rules), then an impression that language acquisition involves a "syntax explosion" is certainly reinforced. This is done in Figure 2.

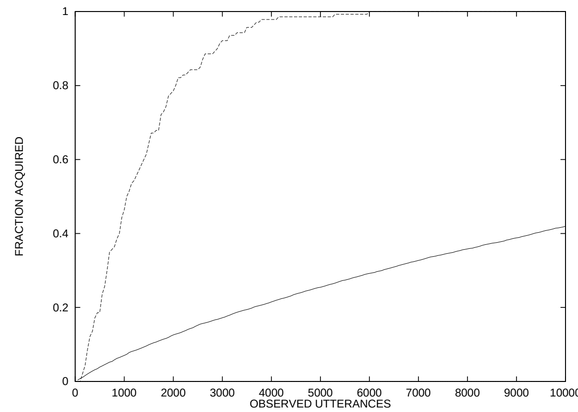

Figure 2. Vocabulary growth by early mechanism, and grammar growth. Lower curve is proportion of entire vocabulary (1000 items) acquired. Upper curve shows proportion of grammar rules (14) acquired. Averaged over 10 runs.

(See a later subsection for another way of quantifying and showing the syntax explosion.)

Up to this point, we have seen two separate mechanisms: an early lexical acquisition mechanism, which paved the way for the semantically driven mechanism involved in the acquisition of grammatical rules. In the next section, I will describe how the acquisition of grammatical rules paves the way, in turn for further, and faster, acquisition of lexical knowledge.

2.6 Later lexical acquisition

As soon as some grammatical rules have been acquired, the meaning of a single new word in an observed sentence can be inferred in one step, provided the context of situation specifies it. With the early lexical acquisition procedure, the learner needed to store many different records of utterances until noticing that several utterances about the same topic contained the same word. But later, given some vocabulary, and grammar rules, the learner can make a partial analysis of an observed utterance that contains a strange word, and if the context of situation provides a concept which none of the known words expresses, it can be safely inferred that the new word means the given concept. This is the same inference as was involved in the last example of the previous section, in which the learner simultaneously acquired a new grammar rule and a new lexical entry. After grammar rules have been acquired, there are far more opportunities than before for acquiring new words and their meanings.

In these simulations, two conditions were compared. In one condition, the grammar-based method of lexical acquisition was not turned on; in the other condition, it was. The growth of vocabulary by the first method only was shown in Figure 1 (only the lower, flatter curve was achieved under relatively realistic Zipfian assumptions about the frequency of concepts in communication). The result of implementing grammar-based lexical acquisition is shown for comparison in Figure 3. The upper curve shows the acceleration of vocabulary growth after some grammar has been acquired.

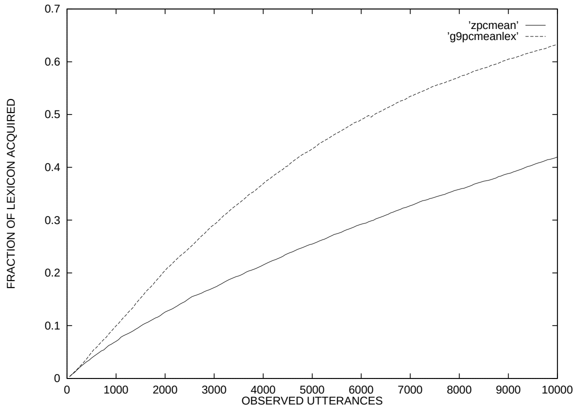

Figure 3. Vocabulary growth by early mechanism (lower curve), and with aid from grammar (upper curve). Curves show proportion of entire vocabulary (1000 items) acquired. Averaged over 10 runs.

Accelerated vocabulary growth leads in turn to faster growth in grammar, as a prerequisite for acquiring a new grammar rule is knowledge of the meaning of most of the words in an utterance. The impressive growth of syntactic competence in young children is hard to quantify. The fundamental observation is that, after about two years of age, children begin rapidly to produce more different types of sentence, longer sentences and more complex sentences. In an attempt to show this overall effect, these simulations surveyed the learner's expressive capacity at all stages during learning. This was done as follows. A random set of 100 meanings was chosen at each measuring stage, and the percentage of these meanings which the child had any way of expressing at all was calculated. This can be thought of as a measure of semantic coverage. The results are shown in Figure 4.

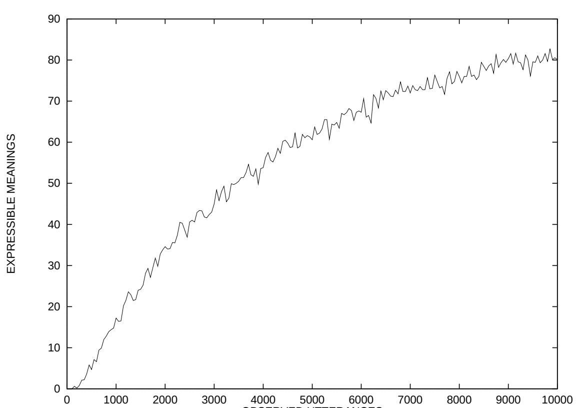

Figure 4. Growth in expressive power, as a percentage of a sample of meanings. Averaged over 10 runs.

2.7 The "stages" are unstaged

The exposition of this paper has been in terms of three phases in language acquisition: early lexical acquisition, grammar acquisition, and later lexical acquisition. But the simulated learner is prepared from the start to follow any of these procedures, whichever is possible at his particular stage of development, for the particular example at hand. The guiding principle is always an attempt to understand, to construct propositions from the observed data. Where the child happens to be able to understand all the words and the syntactic construction in an observed utterance, he will parse it and construct a proposition presumably reflecting the message intended by the speaker. But the same child, at the same (or even a later) stage in life may encounter a totally new sentence with all strange words and an unknown grammatical structure. In this situation he will be forced to fall back on the early lexical acquisition procedure of storing a memory of the utterance and whatever information the context of situation provided, in the hope of later being able to hear enough related sentences to infer the meaning of one of the words involved.

3 Some theoretical context

3.1 Unstressed words and function words

Building in a disposition to treat 'unstressed' words differently, by not looking for conceptual meanings for them, was necessary, as the function words were very frequent, and without this disposition, the child would inevitably conclude that a function word 'meant' some concept that just happened to be part of several propositions expressed by a string with that function word. This treatment of function words seems reasonably well motivated, as it is established (a) that the function words in natural languages are typically unstressed, and (b) that there are significant neurological correlates of the distinction between function words and content words (Tannenhaus, Leiman & Seidenberg 1987; Besner, 1988; Cutler & Morris, 1988; Matthei & Kean, 1989; Shillcock & Bard, 1993). An alternative to building in a disposition to treat function words specially would be a certain kind of purely statistical learning mechanism, powerfil enough to conclude that very frequent words are not significantly correlated with any particular atomic concept. This alternative was not explored here.

3.2 Cross-situational learning

The work reported here overlaps in part with Siskind's (1996) substantial and careful paper. Siskind's simulated learner acquires the meaning of a word by finding "something in common across all observed uses of that word" (41). The present paper applies an essentially similar, but interestingly converse idea. In my simulations, a learner acquires the word for a concept by finding something in common (i.e. a word) across all observed utterances about that concept. The basic idea has been suggested several times in the literature, for example by Pinker (1989) and

Fisher et al (1994). The present paper, like Siskind's is an attempt to model the basic intuitive idea more precisely.

The present work is in several ways complementary to Siskind's. Siskind's model was only concerned with vocabulary acquisition, although he also mentions the contribution that knowledge of grammar can make to this task. Siskind's simulated learners are, unlike mine, given several possible clues from the context of situation about what a word might mean. For Siskind's learner, much of the task is in learning to eliminate possible meanings of words; my simulated learner is given much less information about meanings, but what he is given is (in the simulation) always reliable. Siskind's assumptions about what semantic clues the learner is given are probably closer to reality than mine.

Several previous computational studies have implemented systems in which lexical acquisition is aided by a knowledge of grammar. These include Berwick (1983), Granger(1977) and Jacobs and Zernik (1988).

3.3 Do children learn rules?

This study has treated grammar acquisition as the acquisition of specific grammatical rules for particular constructions, such as actives, passives, dative-shifted constructions, and so on. Clearly, as any linguist knows, the most economical way to state the grammatical facts of a language involves generalization over what have here been treated as individual items of linguistic knowledge, namely grammatical 'rules'. Different linguistic theories have different preferences for the best way of capturing generalizations over the 'rules' of a language. GPSG (Gazdar et al., 1985), for example, uses metarules; the Principles and Parameters (Chomsky, 1981) approach uses parameter settings. I believe it is still an open question whether children actually acquire such compressed representations as are typically sought after by theoretical linguists. There is no space to discuss this complex issue here.

3.4 Degree-0 learnability

Lightfoot (1991) has suggested that a child can build a knowledge of complex grammatical sentences on the basis of triggering experience that only involves simple main clauses. Knowledge of principles of grammatical subordination comes for free. In fact, the simulated learner in the present study also ends up with a knowledge of how to express complex embedded propositions, even though the grammar rules that he has acquired were only ever triggered by exposure to simple (degree-0, non-embedded) sentences. This is because the child is credited with prior knowledge that one proposition can be embedded in another, as an argument of a certain kind of predicate, a predicate of propositional attitude, such as *know* or *believe*. To the learner in this study, the embedded proposition is simply an argument of a predicate which happens to be a proposition.

Having acquired a rule for expressing 2-place predications, such as:

PRED(ARG1, ARG2) ↔ `FORM1 FORM2 FORM3`, where
PRED ↔ `FORM2`
ARG1 ↔ `FORM1`
ARG2 ↔ `FORM3`

the learner is able to treat the condition on the expression of *ARG2* as applying to whole propositions, no less than to individual words. The double arrow in the rule simply means "is mapped by the grammar to". The above rule can have been acquired solely on the basis of simple, degree-0 sentences, like `Bertie loathe Chester`, but, given this rule (and the necessary vocabulary) and other grammar rules pertaining only to simple clauses (such as a rule for passive sentences), the learner in these simulations can now produce complex examples such as the following:

Meaning = *know(max, see(bill, fred))*

```
   max know bill see fred
   max know fred be see -en by bill
   bill see fred be know -en by max
   fred be see -en by bill be know -en by
max
```

In case it is not obvious, the last example is intended to parallel the grammatical, if highly stilted, English sentence "That Fred was seen by Bill was known by Max".

3.5 Critical period for language acquisition

Much has been written about the critical period for language acquisition, from Lenneberg (1967) up to one of the most recent collections (Birdsong, 1999). A brief comment will suffice here.

The simulated learner in the present study does a lot of work in processing each observed utterance early in life. As lexical and grammatical knowledge grows, less work is involved in the processing of individual utterances. Put simply, in the early stages, the child is simultaneously attempting two tasks. He is trying both to understand and to learn language from observed utterances. By the time he has acquired the full lexicon and grammar, no more effort in learning is necessary, and each utterance is simply parsed by the acquired grammar. In the computer program implementing these simulations, the child's first option was always simply to try to parse an input utterance with existing rules. Only if this failed did the child attempt a second strategy of testing to see whether the obstacle to constructing a proposition from the meanings of the known words was a single unknown word or a single unknown grammar rule; if so, the child inferred the necessary lexical entry or grammar rule. If, in turn, this strategy failed, then the child resorted to the brute-force mechanism of storing topic-sentence pairings and subsequently reviewing the stored pairings, as described.

As a human learns her language, it becomes less and less necessary to keep the dedicated language-learning machinery active. From an evolutionary point of view, it would not be surprising if the facility to respond to incomprehensible input withered away after the stage in life when normal language acquisition is complete. (See Hurford, 1991; Hurford and Kirby, 1999, for some detailed discussion of this.)

Acknowledgements

I thank Simon Kirby, Ronnie Cann, and Rob Clark for help with this research. None of it is their fault.

References

R.C. Berwick , "Learning word meanings from examples", *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pp.459-461, Karlsruhe.

D. Besner "Visual word recognition: special purpose mechanisms for the identification of open and closed class items?", *Bulletin of the Psychonomic Society* 26:91-93, 1988.

D. Birdsong , (ed.) *Second Language Acquisition and the Critical Period Hypothesis*, Lawrence Erlbaum Associates, Mahwah, NJ, 1999.

N. Chomsky , *Lectures on Government and Binding: The Pisa Lectures*, Foris Publicatiions, Dordrecht, Holland, 1981

A. Cutler and D.G. Norris, "The role of strong syllables in segmentation for lexical access", *Journal of Experimental Psychology: Human Perception and Performance* 14: 113-121, 1988.

C. Fisher , G. Hall, S. Rakowitz, and L. Gleitman, "When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth", *Lingua*, 92,1:333-375, 1994.

G. Gazdar , E. Klein, G. Pullum and I. Sag, *Generalized Phrase Structure Grammar*, Blackwells, Oxford, 1985.

R.H. Granger , "FOUL-UP: A program that figures out the meanings of words from context", *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pp.172-178, Cambridge, MA.

J.R. Hurford , "The evolution of the critical period for language acquisition", *Cognition*, 40:159-201, 1991.

J.R. Hurford , and S. Kirby, "Co-Evolution of Language-Size and the Critical Period", in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong, Lawrence Erlbaum Associates, Mahwah, NJ, pp.39-63, 1999.

P. Jacobs , and U. Zernik, "Acquiring lexical knowledge from text: A case study", *Proceedings of the Seventh National Conference on Artificial Intelligence*, St Paul, MN, (pp.739-744), 1988.

S. Kirby "Learning, Bottlenecks and the Evolution of Recursive Syntax." to appear in *Linguistic Evolution through Language Acquisition: Formal and Computational Models* edited by Ted Briscoe, Cambridge University Press. (forthcoming)

S. Kirby "Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners", to appear in *The Emergence of Language: Social function and the origins of linguistic form*, edited by C. Knight, M. Studdert-Kennedy and J. Hurford, Cambridge University Press. (forthcoming)

E. Lenneberg , *Biological Foundations of Language*, Wiley, New York, 1967.

D. Lightfoot "The child's trigger experience - degree-0 learnability", *Behavioral and brain sciences*, 14,2:364, 1991.

E.H. Matthei and M-L. Kean, "Postaccess processes in the open vs. closed class distinction', *Brain and Language* 36:163-180, 1989.

S. Pinker , *Language Learnability and Language Development*, Harvard University Press, Cambridge, MA. 1984

R.C. Shillcock , and E.G. Bard, "Modularity and the processing of closed-class words", In G. Altman and R. Shillcock (eds.) *Cognitive Models of Speech Processing* Cambridge, MA.: 163-185, 1993.

J.M. Siskind , "A computational study of cross-situational techniques for learning word-to-meaning mappings" In *Computational Approaches to Language Acquisition*, edited by Michael R. Brent, MIT Press, Cambridge, MA. pp.39-91, 1996.

M.K. Tannenhaus , J.M. Leiman and M.S. Seidenberg "Context effects in lexical processing". In U. Frauenfelder and L. Tyler (eds.) *Spoken Word Recognition*, MIT Press, Cambridge, MA, 1987.