Solving a Lexical Acquisition Task via an Encoding as a Propositional Satisfiability Problem

Jeffrey Mark Siskind
University of Pennsylvania
Institute for Research in Cognitive Science

In the process of learning their native language children must learn a lexicon mapping words to their meanings. In this paper we model a simplified variant of this task as a formal mathematical problem and show how to solve instances of this problem via a reduction to propositional satisfiability. This technique has been implemented and demonstrated effective in solving medium-sized lexical acquisition tasks.

We characterize the lexical acquisition task as follows. The learner is presented with a corpus of utterances, each utterance being a sequence of words. We assume that the learner can hypothesize a set of plausible meanings for each utterance, represented as expressions, from the utterance context given her general perceptual and cognitive abilities. For example, if the learner hears the utterance *Mommy lifted the ball* while observing her mother lifting a ball, we assume that she could form expressions such as $\text{CAUSE}(\textbf{mother}, \text{GO}(\textbf{ball}, \text{UP}))$, $\text{GRASP}(\textbf{mother}, \textbf{ball})$, and $\text{WANT}(\textbf{mother}, \textbf{ball})$ as potential meanings for that utterance. Given such input, the learner faces a two-fold task of first determining which of the hypothesized expressions is in fact the correct meaning of each utterance and then determining which fragments of that expression denote the meanings of the individual words in that utterance. In the above example the learner must first determine that $\text{CAUSE}(\textbf{mother}, \text{GO}(\textbf{ball}, \text{UP}))$ is the correct meaning of *Mommy lifted the ball* and then determine that **mother**, $\text{CAUSE}(x, \text{GO}(y, \text{UP}))$, and **ball** are the correct meanings of *Mommy*, *lift*, and *ball* respectively and not vice versa. Siskind (1993) gives a more detailed definition of this problem.

The key insight behind our approach is the observation that given the particular linking rule adopted in Siskind (1993):

- each functor in the expression representing the meaning of an utterance must be contributed by exactly one word in that utterance, and
- all functors contained in the meanings of all words appearing in an utterance must appear in the meaning of that utterance.

For example, if by some process the learner could determine that *John walked to school* meant $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$ and could also determine that *school* meant **school** then she could infer that *to* could not mean $\text{TO}(\textbf{school})$ since the symbol **school** appears only once in the expression $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$ and that symbol is already contributed by the word *school* and thus could not be contributed a second time by another word such as *to*. Similarly, if the learner could determine that *John walked to school* meant $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$, that *John* meant **John**, that *to* meant $\text{TO}(x)$, and that *school* meant **school**, then she could infer that the expression representing the meaning of *walked* must contain the functor GO since that symbol appears in $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$ but is not contributed by any of the words *John*, *to*, or *school*. Finally, if the learner could determine that *John walked to school* meant $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$ then she could rule out **Mary** as a possible meaning for *John* since while the word *John* appears in *John walked to school*, the symbol **Mary** does not appear in $\text{GO}(\textbf{John}, \text{TO}(\textbf{school}))$.

Our system computes the aforementioned inferences by first translating the input corpus into a set of propositional formulas and then solving the resulting propositional satisfiability problem For example, we represent the statement that the word *John* means **John** by the conjunction $\neg\bot_{John} \wedge p_{John,\textbf{John}} \wedge \neg p_{John,\textbf{Mary}} \wedge \cdots$ and the statement that the word *lift* means $\text{CAUSE}(x, \text{GO}(y, \text{UP}))$ by the conjunction $\neg\bot_{lift} \wedge p_{lift,\text{CAUSE}} \wedge p_{lift,\text{GO}} \wedge p_{lift,\text{UP}} \wedge \neg p_{lift,\textbf{John}} \wedge \cdots$. Here propositions of the form p_{wf} denote the statement that the symbol f appears in the expression denoting the meaning of the word w while propositions of the form $\bot_w$ denote the statement that the word w is devoid of any semantic content and thus does not contain any symbols in the representation of its meaning. Examples of such words are case markers and complementizers.

The process of translating the input corpus into a set of propositional formulas involves instantiating a number of axiom schemas over the propositional variables p_{wf} and $\bot_w$. One such axiom schema states that a word cannot both be devoid of any semantics and also contribute a functor.

$$\bigwedge_{w\in\bigcup_{i=1}^{n} u_i} \;\; \bigwedge_{f\in\bigcup_{i=1}^{n}\bigcup_{k=1}^{l(i)} F(t_{ik})} \neg(\bot_w \wedge p_{wf})$$

Here n denotes the number of utterances in the corpus, $l(i)$ denotes the number of different meanings hypothesized for utterance u_i, and $F(t_{ik})$ denotes the set of all functors appearing in the expression denoting the k^{th} meaning associated with utterance u_i. This axiom allows one to infer that if the word *of* is devoid of any semantics then the functor GO cannot be part of its meaning. In its contrapositive form it allows one to infer that if the word *lift* contains GO as part of its meaning then it cannot be devoid of any semantics. Another axiom schema states that each functor in the correct meaning of an utterance must be contributed by at least one word of that utterance.

$$\bigwedge_{i=1}^{n}\bigwedge_{k=1}^{l(i)}\bigwedge_{f\in F(t_{ik})}\left(q_{ik} \rightarrow \bigvee_{w\in u_i} p_{wf}\right)$$

Here q_{ik} denotes the statement that t_{ik} is the correct meaning of u_i. This axiom allows one to infer that the meaning of the word *saw* must contain the function SEE, if *John saw Mary* meant $\text{SEE}(\textbf{John}, \textbf{Mary})$ and neither *John* nor *Mary* contained the functor SEE in their meaning. In its contrapositive form it allows one to infer that *John saw Mary* could not mean $\text{SEE}(\textbf{John}, \textbf{Mary})$ if none of the words *John*, *saw*, or *Mary* contained the functor SEE in their meaning. Our system incorporates nine such axiom schemas.

We have applied our technique to a number of randomly generated corpora of English-like and Japanese-like text. Siskind (1993) describes the techniques used to generate such corpora. The results are summarized in the following table:

	'English'			'Japanese'	
Utterances	9	200	2000	9	200
Words in lexicon	11	102	992	11	105
Words in corpus	32	955	9489	39	1122
CPU Sec (SPARCstation ELC)	56	63	919	47	55

Our technique can quickly learn a lexicon of 1000 word-to-meaning mappings from a corpus of 2000 utterances ambiguously paired with four possible meanings. We believe that our technique will scale up by an order of magnitude to solve tasks of the size faced by real children. While we do not claim that children actually learn language using our technique, we do believe that they could—in principle—do so.

A key characteristic of our method is that it requires no knowledge or use of syntax on the part of the learner. Utterances in the corpus are treated as unordered multisets of words. This has important implications for the semantic-syntactic bootstrapping debate. Proponents of syntactic bootstrapping argue that semantic bootstrapping presupposes cross-situational learning and that cross-situational learning is too weak to succeed. Our results formally characterize the weakness of cross-situational learning as previously conceived, yet show how the new method forms a stronger technique which operates using exactly the same input data. This severely weakens that argument in favor of syntactic bootstrapping and reopens the possibility that substantially empiricist methods can be used first to acquire word-to-meaning mappings and then acquire syntax using knowledge of the lexicon previously acquired.

References

Jeffrey Mark Siskind. Lexical acquisition as constraint satisfaction. Submitted to ACL93.