SEMANTIC BOOTSTRAPPING OF GRAMMAR IN EMBODIED ROBOTS

YO SATO
Adaptive Systems Group, University of Hertfordshire
College Lane, Hatfield, AL10 9AB, U.K.
y.sato@herts.ac.uk

JOE SAUNDERS
Adaptive Systems Group, University of Hertfordshire
College Lane, Hatfield, AL10 9AB, U.K.
j.1.saunders@herts.ac.uk

1. Situated grammar learning from partially word-detected string

The concept of semantic bootstrapping of grammar, though tracing back at least to Pinker (1984), has not been explored fully in applied computational linguistics as a grammar learning algorithm. Experimental methods with embodied robots may however open up the possibility to exploit their fine sensori-motor capabilities to examine the feasibility of such an algorithm. We report one such attempt.

In view of the difficulty of word segmentation from audio stream, we start from the stage where only *some* of the constituent words have been detected in an utterance, marking a departure from most grammar induction methods, semantically oriented or not, which presuppose the full set of terminal symbols.

Such partially word-detected strings are obtained from the human-robot interaction experiments, in which the human participant is asked to explain simple objects verbally to the robot. The recorded speech is first converted with a speech recogniser to phoneme strings. We then create a small vocabulary of the frequently ocurring words and convert only those phoneme sequences that have a match in this vocabulary into words, creating what we call a phoneme-word list (PWL). For example, if our vocabulary list is ⟨sɔːkl (‘*circle*’), rɛd (‘*red*’), siː (‘*see*’), juː (‘*you*’)⟩, then the phoneme string ⟨juːsiːðætrɛdstaːʃeɪp⟩ (‘You see that red star shape’) is converted into ⟨[*you*], [*see*], ðæt, [*red*], staːʃeɪp⟩, where the elements in square brackets represent ‘detected’ words.

Another crucial piece of data is a possible meaning relevant to a given situation, i.e. meaning hypothesis. Thus, the proposal amounts to the learning of syntax from the pairs of a meaning hypothesis and a PWL, e.g.:

$$see'(r,o) \wedge red'(o) \wedge star'(o) \quad : \quad$$ ⟨[*you*], [*see*], ðæt, [*red*], staːʃeɪp⟩

where *r* and *o* represent the individual constants for the robot learner and the object talked about. Notice that the semantics is not fully spelt out: the predicate corresponding to *shape* is missing. We build consistent *grammar* hypotheses such an underspecified hypothesis in a manner to be described below.

2. Grammar learning as lexical type induction

Taking inspiration from Fulop's (2004) type-logical work on samantic grammar induction in a lexicalist but broadly Montagovian framework (Montague, 1973), we characterise grammar learning as the induction of syntactic behaviour of words from their *semantic types* that could compose a meaning hypothesis. However we employ a typed-feature structure (TFS) grammar in a manner of HPSG (Pollard & Sag, 1994), taking advantage of its underspecifiability inherent in unification, so that the intermediate stages of learning can be represented succinctly. Our PWL can be characterised as a list of underspecified TFSs, as below:

$$\left\langle \begin{bmatrix} \text{word} \\ \text{PHNMS} \begin{bmatrix} \text{phnms} \\ \text{juː} \end{bmatrix} \\ \text{SEM} \mid \text{TY ??} \\ \text{SYN ??} \end{bmatrix}, \begin{bmatrix} \text{word} \\ \text{PHNMS} \begin{bmatrix} \text{phnms} \\ \text{siː} \end{bmatrix} \\ \text{SEM} \mid \text{TY ??} \\ \text{SYN ??} \end{bmatrix}, \begin{bmatrix} \text{phnms} \\ \text{ðæt} \end{bmatrix}, \begin{bmatrix} \text{word} \\ \text{PHNMS} \begin{bmatrix} \text{phnms} \\ \text{rɛd} \end{bmatrix} \\ \text{SEM} \mid \text{TY ??} \\ \text{SYN ??} \end{bmatrix}, \begin{bmatrix} \text{phnms} \\ \text{stɑːʃeɪp} \end{bmatrix} \right\rangle \quad \Bigg| \quad \text{the first word becomes:} \begin{bmatrix} \text{word} \\ \text{PHNMS} \begin{bmatrix} \text{phnms} \\ \text{juː} \end{bmatrix} \\ \text{SEM} \mid \text{TY e} \\ \text{SYN} \mid \text{LP} \prec \text{et} \end{bmatrix}$$

This PWL can be taken as a compact representation of *any* string that satisfies the given constraints. The grammar induction, then, boils down to populating the as yet empty features (with '??') for the SEM(ANTICS) and SYN(TAX) features.

As semantic bootstrapping, the immediate target of learning is the semantic TY(PE) feature. If, for example, the TY value of the first word, *you*, is hypothesised to the type *e* (individual) (see the right hand side in the above figure), then the hypothesising on the TY feature of the other words as well as the SYN feature can start: given the assumption that the utterance is a truth-value bearing type, *t*, one can predict that one of the following words is part of the *et* type, and that our present word, *you*, may be hypothesised to linearly precede the *et* type.

3. Preliminary results and future tasks

The preliminary evaluations of the initial experiment indicate that the size of the hypothesis space may be problematic. For a relatively quick convergence of hypotheses, it would need to be reduced. Possible methods include cognitively plausible biases and, perhaps more contentiously, corrective feedback from humans.

References

Fulop, S. A. (2004). Semantic bootstrapping of type-logical grammar. *Journal of Logic, Language and Information*, *14*, 49-86.

Montague, R. (1973). The proper treatment of quantification in ordinary English. In Hintikka (Ed.), *Approaches to natural language*.

Pinker, S. (1984). *Language learnability and language development*. Harvard UP.

Pollard, C., & Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. CSLI.

The work described in this paper was conducted within the EU Integrated Project ITalk ("Integration and Transfer of Action and Language in Robots") funded by the European Commission under contract number FP7-214668.