3B-01

# A Learning Model of Lexicon

Sumiyo Nishiguchi

Department of Management, Tokyo University of Science

**Abstract**

In Gold's (1967) model, a learner is a function from information to guesses on grammar at each time. This paper models a learning process of lexical knowledge and considers a learner to be a function from a set of lexicon to a poset of lexical entries. A polysemous word may be assigned a set of entries while uncertainty maps a word to multiple candidates in Boolean algebra. The lattice of lexical entries has a null element in case no meaning is assigned.

## 1. Gold (1967)

In Gold's (1967) model, a learner is a function from information to guesses on grammar at each time.

Gold (1967) defines a method of information presentation in learning processes of language as in (1).

(1) $g_t = G(i_1, \cdots i_t)$

Quantized time starts at 1.

(2) $t = 1, 2, \cdots$

At each time, a learner G makes a guess $g_t$ of a name in an unknown language L based on the information at t, $i_t$.

## 2. Learning Lexicon

For learning processes of lexicon, I assume that the power set of the set of lexicon L is returned upon acquisition of the units of information.

(3) $G(i_1, \cdots, i_t) = \wp(L)$

It is because there could be several guesses in the middle of learning stage. A learner may not be sure whether the sound ``dog" refers to a dog or cat as shown in (4). On the other hand, even after mastering a language, a word can be polysemous: e.g., *bat* means both a bird or a baseball bat as om (6).

(4) G(/dog/) = {dog, cat}

(5) G(/bæt/) = $\{bat_1, bat_2\}$

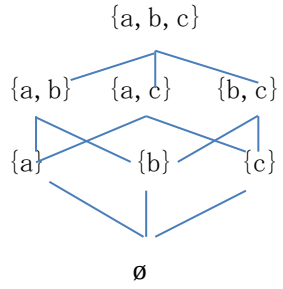
In fact, the learner returns either conjunction for lexical ambiguity or disjunction for uncertainty.

(6) $G(/bæt/) = bat_1 \wedge bat_2$

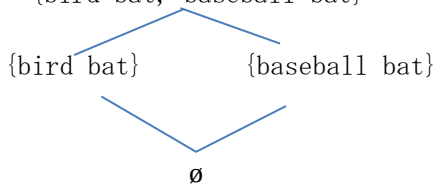
(7) $G(/dog/) = dog \vee cat$

The returned values are the partially ordered set or poset $\langle L, \leq \rangle$ which is a binary relation over a set L which is reflexive, antisymmetric, and transitive

(8) The Hasse diagram of $\wp(L)$ where L = {a, b, c}

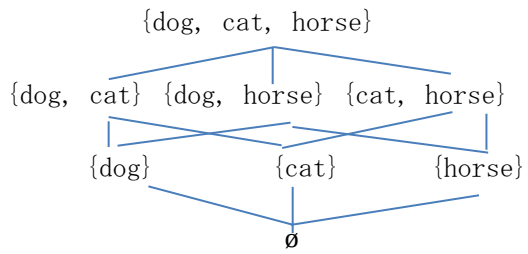



(9) Lexical Ambiguity: G(/bat/)




When the learner does not know the referent of /dog/, there could be three candidates as a referent:

(10) Uncertainty: G(/dog/)




Considering lexical ambiguity as meet and uncertainty as joint, meet a ∧ b and the join a ∨ b of two elements coincide with their infimum and supremum, respectively, with respect to ≤.

## 3. Further

As a recent study on a learning model of lexicon, Elsener et al. (2013) discusses Baysian model. The present study models fundamental insight on lexical ambiguity and uncertainty.

In Gold's model, after conversion, or completion of learning a language, the learner returns the same guesses. However, the learning process of lexicon does not seem to converge. The number of vocabulary may increase after learning grammar – a person may learn a new word created every year, or forget some words.

## References

Elsner, Micha, Sharon Goldwater, Naomi Feldman, and Frank Wood (2013) "A Joint Learning Model of Word Segmentation, Lexical Acquisition, and Phonetic Variability, " *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Gold, E. Mark (1967) "Language Identification in the Limit," *Information and Control* 10, 447-474.