The Significance of Errors to Parametric Models of Language Acquisition

Paula Buttery
Natural Language and Information Processing Group
Computer Laboratory, Cambridge University
paula.buttery@cl.cam.ac.uk

Classification of Language Examples

Children become fluent despite lack of formal language teaching.

Not every utterance heard is a valid example of the environment language.

How can the child know which utterances are valid?

Every time a child mis-classifies an utterance as valid we get an error.

Sources of Error

- Accidental Errors: lapses of concentration, slips-of-the-tongue, interruptions.
- Ambiguous Environments: bi-lingual environments, diglossia, language change
- Indeterminacy of Language:
  - Indeterminacy of meaning: "John kissed Kate" vs. "Kate was kissed by John"
  - Indeterminacy of parameter settings: SVO vs. SOV with v2

Require a learning model to attempt to learn from every utterance and be unaffected by misclassification errors.

The Numbers Game

Game with 2 players:

- Player One: thinks of a set of numbers that can be defined by a rule.
- Player Two: attempts to discover the rule defining the set.

Only information available to player two is a stream of examples from player one.

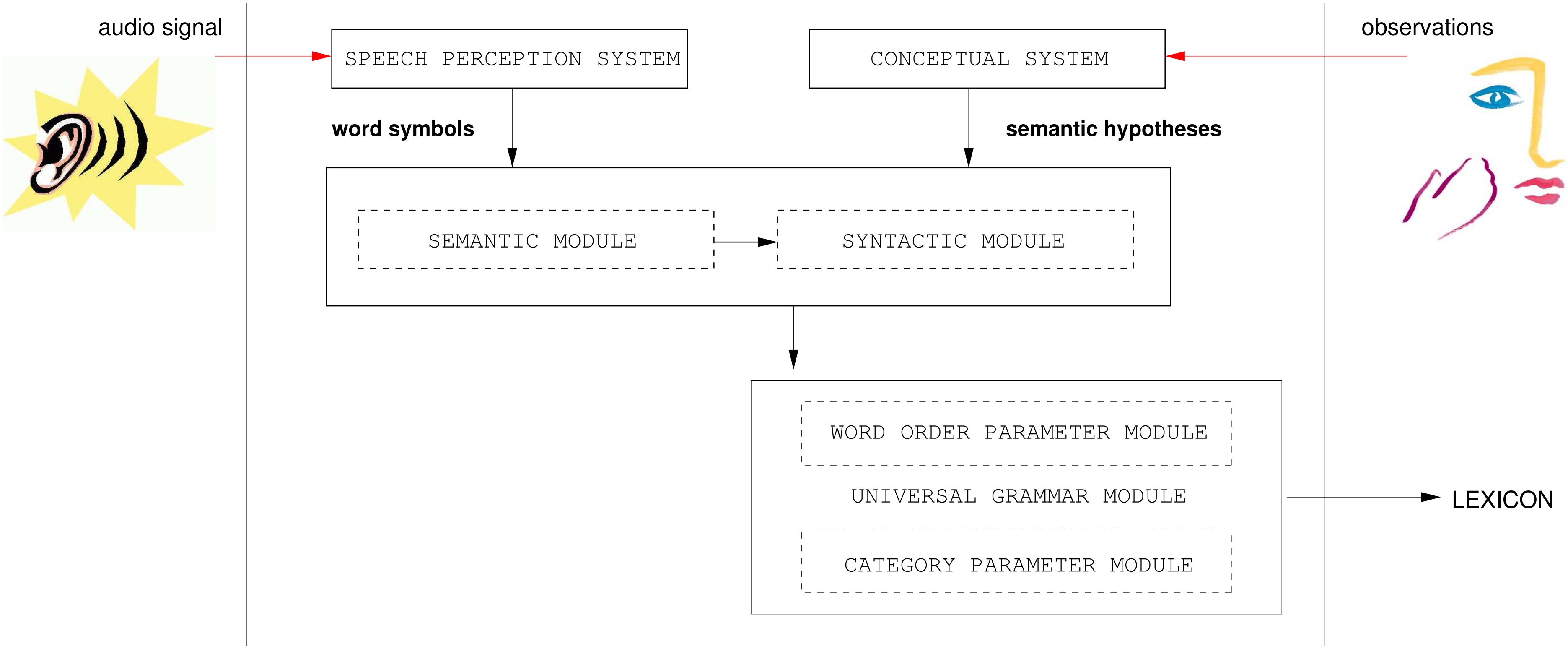
Deterministic Learners

Gibson and Wexler's Trigger Learner:

- Algorithm:
  - attempt to parse with current parameters;
  - change one parameter;
  - adopt new settings if we can analyze an utterance that was previously not analyzable.
- Problems:
  - local maxima;
  - worse case scenario - last utterance seen is an error.

Gibson E and Wexler K, 1994. Triggers. *Linguistic Inquiry* 25(3): 407-454

A Robust Learning System



Semantics Learning Module

Cross Situational Techniques:

- Constraining Hypotheses with Partial Knowledge:

  If learner knows that: “cheese” $\mapsto$ **cheese**

  and on hearing “Mice like cheese” hypotheses:

  like(mice, cheese)

  madeOf(moon, cheese)

  madeOf(moon, cake)

  then we can rule out **madeOf(moon, cake)**

Siskind J. 1996. A computational study of cross situational techniques for learning word-to-meaning mappings. *Cognition* 61(1-2):39-91

Syntactic Learning Module

Hypothesizes categorial grammar categories for a word:

- Forward Application ($>$)

  $X/Y \quad Y \rightarrow X$

- Backward Application ($<$)

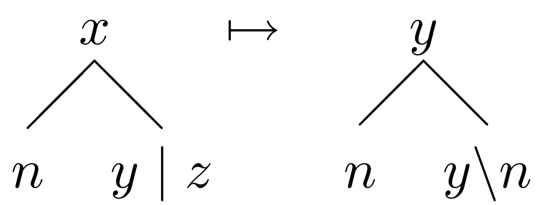
  $Y \quad X\backslash Y \rightarrow X$

$$\dfrac{\dfrac{Kim}{np} \qquad \dfrac{\dfrac{likes}{(s\backslash np)/np} \quad \dfrac{Sandy}{np}}{s\backslash np}>}{s}<$$

Syntactic Learning Module

Typing Assumption: the semantic arity of a word is usually the same as its number of syntactic arguments.

verb(arg1 ,arg2) $\mapsto a \mid b \mid c$



The Universal Grammar

Underspecified inheritance hierarchy:

- Categorial Parameters: 60 parameters
  - one per legal syntactic category
- Word Order Parameters: 18 parameters
  - e.g. subject direction parameter (SVO,SOV vs. OVS,VSO)

Universal Grammar module consulted whenever syntactic learner returns a valid syntactic category for every word.

The Sachs Corpus

Natural interactions of a child with her parents:

- Real child-directed utterances - child's utterances removed;
- Corpus modeled by Villavicencio;
- Annotated with semantic representations.

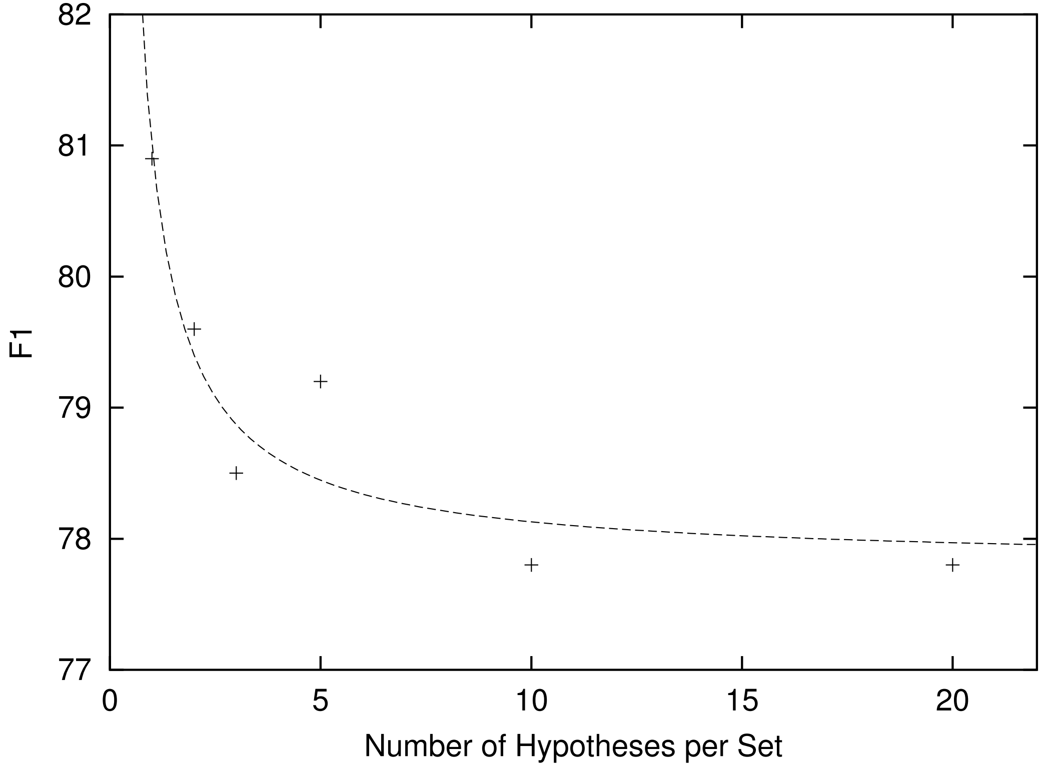
Villavicencio A. 2002. *The acquisition of a unification based generalized categorial grammar* Ph.D Thesis, University of Cambridge.

Exp. 1: Indeterminacy of Meaning

Increasing numbers of semantic hypotheses per utterance:

- Extra hypotheses chosen randomly.
- Correct semantic expression was always present in the set.
- Hypothesis sets of sizes 2, 3, 5, 10 and 20.

Exp. 1: Indeterminacy of Meaning



Exp. 2: Indeterminacy of Parameter Settings

Misclassification due to thematic role: “He likes fish”

Possible interpretations:

likes(he, fish) - SVO

likes(fish, he) - OVS

- Learner was exposed to increasing amounts of misinterpreted thematic role (0% to 50% of all occurances)

Exp. 2: Indeterminacy of Parameter Settings

- mis-classification varied between 0% and 50% at 10% intervals:
  - 9 word-order-parameters set;
  - 13.5 word-order-parameters correct according to target (due to inheritance).
  - 45% difference in speed of convergence between error-free and maximum thematic-role-error case.

Conclusions

Errors due to misclassification of language examples are likely.

Deterministic parametric learners have problems handling errors.

A statistical error-handling learner may be robust to errors.

Indeterminacy of language is just another case of misclassification.

Natural Language and Information Processing Group: *www.cl.cam.ac.uk/users/ejb/*

email to: paula.buttery@cl.cam.ac.uk.