Language and Computation

week 11, Thursday, April 10

Tamás Biró
Yale University
tamas.biro@yale.edu
http://www.birot.hu/courses/2014-LC/

Practical matters

- **Post-reading:** JM 11, 17, 20.1-3.
- **Pre-reading:** JM 18.1, 19.1, 20.1, 21.1.
- `http://birot.hu/courses/2014-LC/readings.txt`
- Assignment 4 due. Assignment 5 to come soon.
- Midterm returned.

Today

- Stochastic approaches to Optimality Theory
- Optimality Theory: learnability
- Introduction to semantics
- Selected topics in computational semantics

Next time: Computational discourse and dialogue systems.

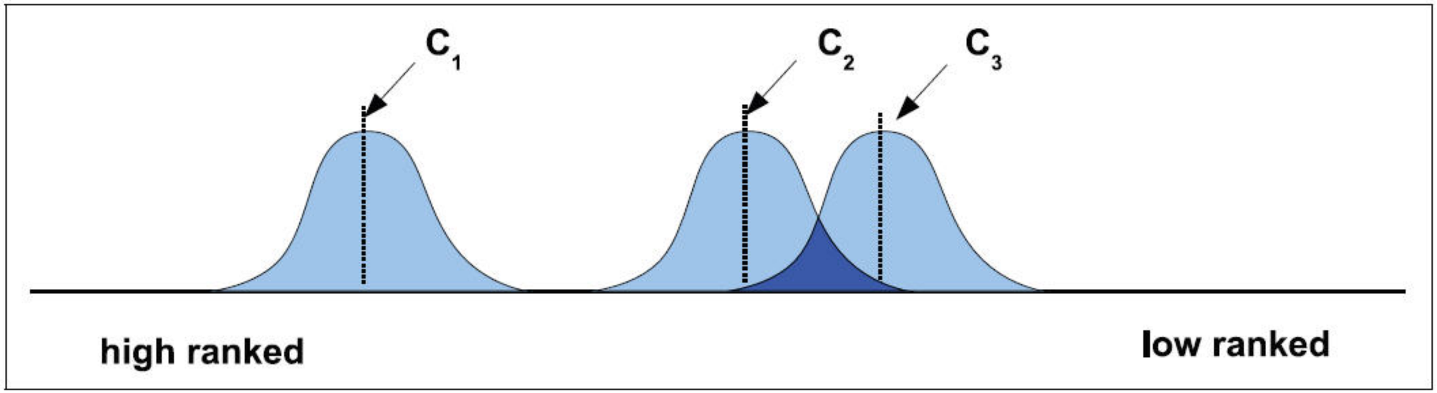
Stochastic approaches to Optimality Theory

Stochastic grammars: why?

- Frequency in corpora? — No! (Or yes?)
  (*I was born in New Haven* vs. *I was born in New York*)
- Free variation: more than one grammatical form
  . . . being produced by a single brain
  . . . being produced by speakers of a language community
  (*more stupid* vs. *stupider*)
- Gradient grammaticality judgement
- Performance errors (e.g., fast speech errors)

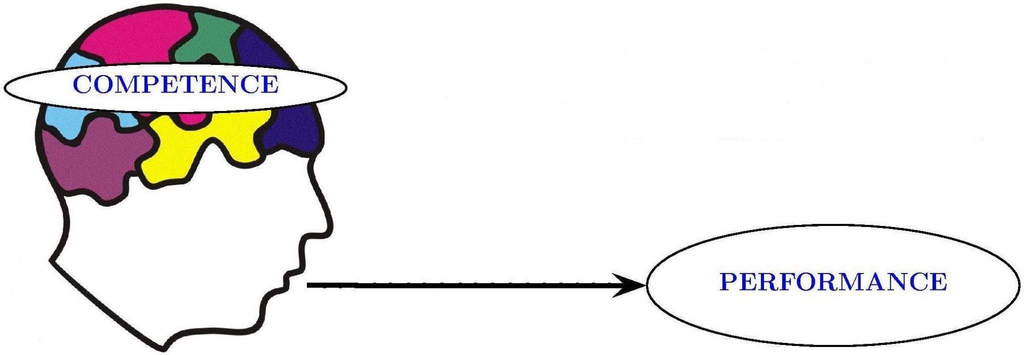
Variation in Optimality Theory

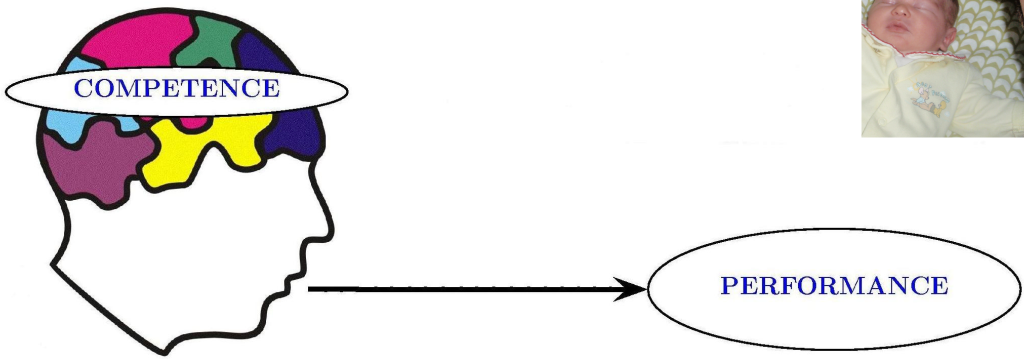
- More elements in $\mathrm{Gen}(U)$ with same violation profile.
- Implementation can return other candidates than the (globally) optimal element of $\mathrm{Gen}(U)$.
- 1 mental grammar $=$ stochastic combination of more "elementary grammars". E.g, Boersma's *Stochastic OT*:



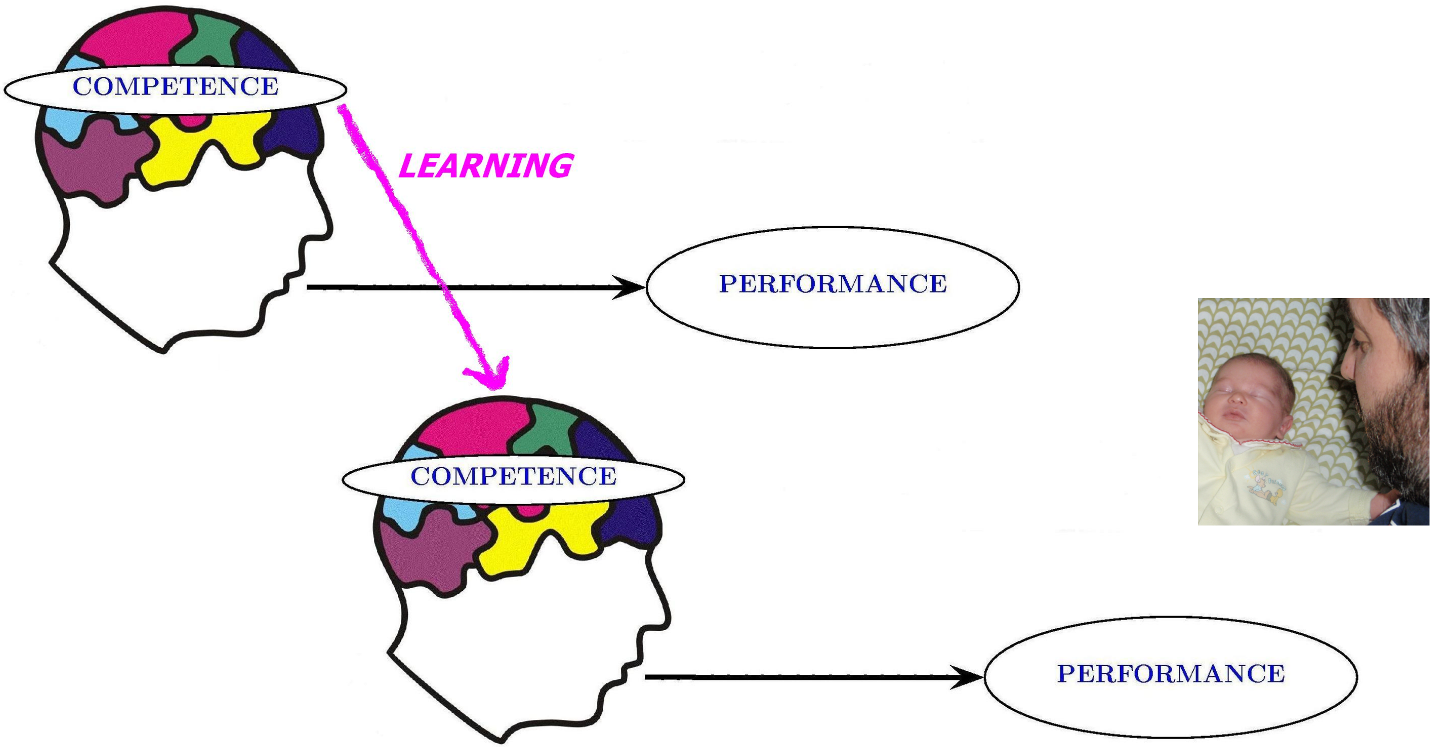
Learning Optimality Theory

Language acquisition

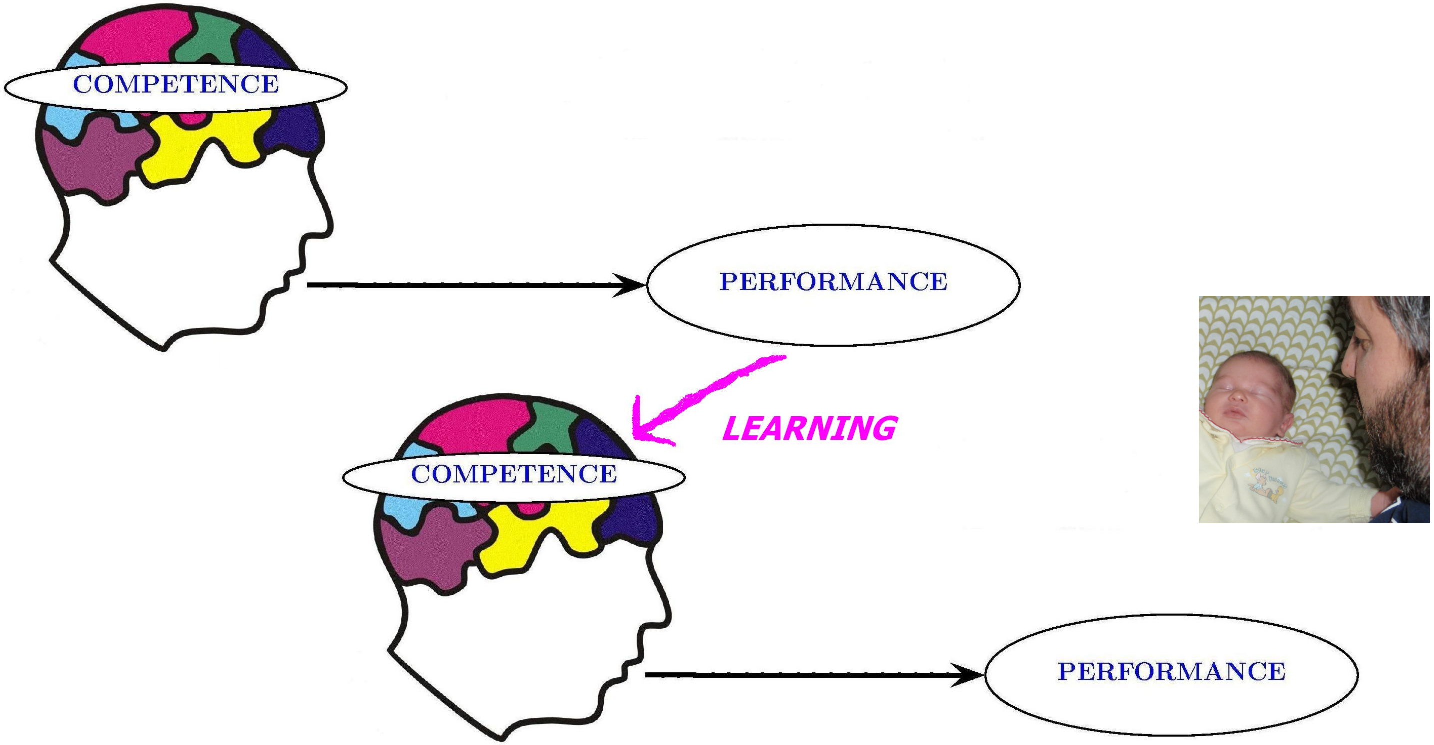




Language acquisition



Language acquisition



Learning in Optimality Theory

General idea:

- Speaker-teacher wants to say *underlying form* uf.
- Speaker-teacher's grammar produces *surface form* sf.
- Listener-learner hears *surface form* sf = *winner form* w.
- Listener-learner's grammar would produce uf as *loser form* l.
- Listener-learner updates her grammar, in order to produce w, and not l:

  Winner-preferring constraints are promoted and loser-preferring constraints are demoted in hierarchy hypothesized by the learner.

Learning in Optimality Theory

General idea:

	/underlying form/	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
	Candidate 1 (learning observation)	$*!\longrightarrow$	$*\longrightarrow$			$*\longrightarrow$			
☞	Candidate 2 (learner's output)				$\longleftarrow*$		$\longleftarrow*$		

- Winner preferring constraints vs. Loser preferring constraints
- All L must be dominated by at least one W.
- Demote L, possibly promote W.

Learning in Optimality Theory

General idea:

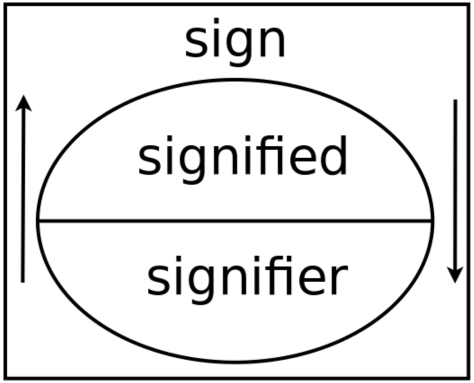
/underlying form/	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
Candidate 1 (learning observation)	$*!\longrightarrow$	$*\longrightarrow$			$*\longrightarrow$			
☞ Candidate 2 (learner's output)				$\longleftarrow*$		$\longleftarrow*$		

- Recursive Constraint Demotion: off-line (batch learning)
- Error Driven Constraint Demotion: on-line
- Gradual Learning Algorithm

Semantics

Semantics: the study of 'meaning'

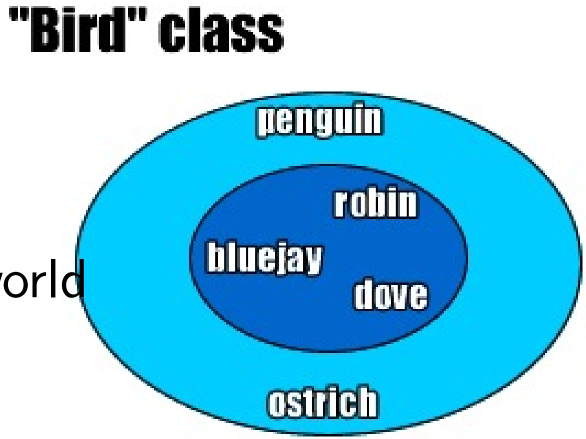
F. de Saussure (1916): linguistic sign



signifier:	phonetics, phonology,	morphology, syntax
signified:	–	semantics

What is meaning?

- A mental construct? Category formation:
  - Prototypes
  - Examplars



- Reference theories: what the linguistic sign refers to in the world
  '*The current king of France*'?

- Truth value:
  The set of possible worlds in which the proposition holds.

What is meaning?

- Lexical semantics: "atomic units"
- Compositional semantics:
  from atomic units to the meaning of phrases and sentences.

What is meaning?

- WE DO NOT KNOW IT!

- But let us handle it. . .

- How to do it?

Why handle meaning?

- Seemingly,
  most "ultimate" NLP tasks require access to meaning:

  machine translation, question answering, information extraction, dialogue systems, spell checking, etc.

  at least, when we think of the way humans solve these tasks.

- To improve quality of "lower level" NLP tasks:

  speech synthesis and recognition, part-of-speech tagging, morphological and syntactic parsing, etc.

How to handle meaning?

- By tackling the problem:
  Create a computational model of the mental representation of the world. . . Hope to do so in the 60s, but then given up.

- By circumventing the problem:
  E.g., Probabilistic Grammars with corpus based frequencies.

- By employing intermediate solutions

Word sense

- Create a computational model of the mental representation of the world. . . Hope to do so in the 60s, but then given up.

- Its usasge: a vector of contexts in which the word is used in the corpus.

$\rightarrow$ WDS: word sense disambiguation,
a classic example of Machine Learning.

See you next week!