

The Verb

From Probabilities to Internal Categories

Cem Bozsahin

bozsahin@metu.edu.tr

Cognitive Science Department, The Informatics Institute, ODTÜ

November 18, 2016

Joint work with Mark Steedman
(with thanks to Luke Zettlemoyer)

To Appear as:

Steedman and Bozsahin (2017). Projecting from the Lexicon, submitted
Bozsahin (2017). Verbs, Heads, Word Orders and Types. in *Word Order in Turkish*, Springer, Ozsoy (ed.).

- Thematic structure
- Argument structure
- Word order
- Tense
- Aspect
- Mood

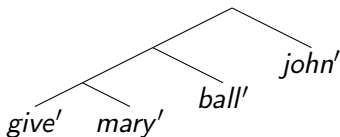
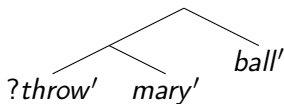
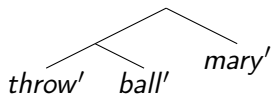
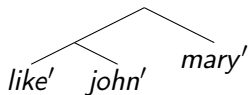
They have a lot to do with how we report **ACTIONS** and **EVENTS**.

A brief history of Linguistics

- Thematic structure from argument structure
- Argument structure from thematic structure
- How do we get syntactic forms, logical forms and phonological forms ON THE FLY?
- These “interfaces” must be quite trivial

- There seems to be a universal conceptual space (things, actions, events)
- That space seems to be asymmetric
- All human languages are probably homomorphic in that space
- Since languages differ in surface structure, surface structure cannot be homomorphic to that space
- That brings us to WORD ORDER
- How do we capture BOTH aspects if parsing is a reflex? (Garrett)

Asymmetry and structure



$(((\text{give}' \text{ mary}') \text{ ball}') \text{ john'}) =$
 $\text{give}' \text{ mary}' \text{ ball}' \text{ john}'$

(1) a. The window broke.

$init'(broken'w')$

b. The stone broke the window.

$\lambda z.cause'(init'(broken'w'))st'z$

c. The man broke the window with a stone.

$cause'(init'(broken'w'))st'man'$

$init'state'$: an inchoative event culminating in state $state'$

What are these primes and LFs? Hypotheses about LOT.

- (2) a. $\text{persuades} := (S \setminus NP_{3s}) / VP_{\text{to-inf}} / NP$ English
- b. $\text{promises} := (S \setminus NP_{3s}) / VP_{\text{to-inf}} / NP$
- c. $\text{expects} := (S \setminus NP_{3s}) / VP_{\text{to-inf}} / NP$
- d. $\text{broke} := (S \setminus NP) / PP / NP$
- e. $\text{broke} := (S \setminus NP) / NP$
- f. $\text{broke} := S \setminus NP$
- g. $\text{gwelodd (saw.3s)} := (S / NP) / NP_{3s}$ Welsh
- h. $\text{iniutusan (order)} := (S_{\text{DV}} / VP) / NP_{\text{ANG,AGR}} / NP_{\text{NG}}$ Tagalog
- i. $\text{savundum (defended.1s)} := S \setminus S'_{\text{acc}}$ Turkish
- j. $\text{viu (saw.3s)} := (S \setminus NP_{3s}) / NP$ Portuguese

- (3) a. The window broke.

$\text{broke} := S \backslash NP : \lambda x. \text{init}'(\text{broken}'x)$

- b. The stone broke the window.

$\text{broke} := (S \backslash NP) / NP : \lambda x \lambda y \lambda z. \text{cause}'(\text{init}'(\text{broken}'x))y z$

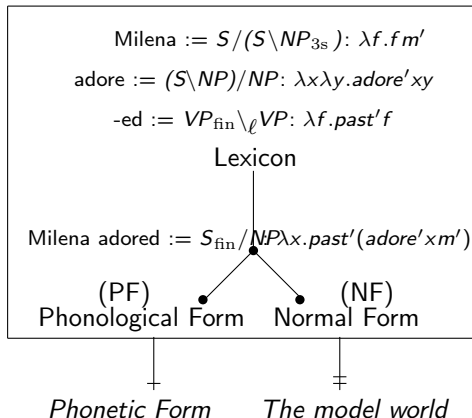
- c. The man broke the window with a stone.

$(S \backslash NP) / PP / NP : \lambda x \lambda y \lambda z. \text{cause}'(\text{init}'(\text{broken}'x))y z$

We get thematic structure as an inference over predicate-argument structure

If surface structure is not homomorphic to universal conceptual structure, how do children learn the word order of their language?

Linguistic architecture



- combinatory projection to constituents of *string* := *syn:sem*
- serialization of feature geometry from *string* and *syn*
- normalization from *syn* and *sem*
- + realization and intake
- ≠ inference and valuation

- Syntax is a HIDDEN variable for the child.
 - She has to infer a CATEGORY from form-meaning pairs only.
 - These pairs are subject to universal combinatorics. (PARSING)
 - These inferences set up a MENTAL GRAMMAR.
 - Such grammars are necessarily PROBABILISTIC.
-
- Change the nature of categories, and we might end up with PLANS and SCRIPTS.

Now with full categories

- (4) a. persuades :=
 $(S \setminus NP_{3s}) / VP_{to-inf} / NP: \lambda x \lambda p \lambda y. persuade'(px)xy$
- b. promises :=
 $(S \setminus NP_{3s}) / VP_{to-inf} / NP: \lambda x \lambda p \lambda y. promise'(py)xy$
- c. expects := $(S \setminus NP_{3s}) / VP_{to-inf} / NP: \lambda x \lambda p \lambda y. expect'(px)y$
- d. broke := $(S \setminus NP) / PP / NP: \lambda x \lambda y \lambda z. cause'(init' broken' x)yz$
- e. broke := $(S \setminus NP) / NP: \lambda x \lambda y \lambda z. cause'(init' broken' x)yz$
- f. broke := $S \setminus NP: \lambda x. init'(broken' x)$
- g. gwelodd (saw.3s) := $(S / NP) / NP_{3s}: \lambda x \lambda y. saw'yx$
- h. iniutusan (order) :=
 $(S_{DV} / VP) / NP_{ANG,AGR} / NP_{NG}: \lambda x \lambda y \lambda p. order'(py)yx$
- i. savundum (defended.1s) :=
 $S_{pd} \setminus S'_{acc}: \lambda p. defend'pi' \wedge topic'i'$
- j. viu (saw.3s) := $(S \setminus NP_{3s}) / NP_{rex}: \lambda x \lambda y. saw'xy$

Categorical Word order: not a metrical concept

- (5) a. $(S \backslash NP) / NP: \lambda x \lambda y. verb' xy$ (SVO)
b. $(S / NP) \backslash NP: \lambda x \lambda y. verb' yx$ (SVO')
c. $(S / NP) \backslash NP: \lambda x \lambda y. verb' xy$ (OVS)
d. $(S \backslash NP) / NP: \lambda x \lambda y. verb' yx$ (OVS')
e. $(S \backslash NP) \backslash NP: \lambda x \lambda y. verb' xy$ (SOV)
f. $(S \backslash NP) \backslash NP: \lambda x \lambda y. verb' yx$ (OSV)
g. $(S / NP) / NP: \lambda x \lambda y. verb' xy$ (VOS)
h. $(S / NP) / NP: \lambda x \lambda y. verb' yx$ (VSO)

Categorical Word order: not a metrical concept

(6) a.	$(S \backslash NP) / NP: \lambda x \lambda y.verb'xy$	(SVO)	English
b.	$(S / NP) \backslash NP: \lambda x \lambda y.verb'yx$	(SVO')	Huastec
c.	$(S / NP) \backslash NP: \lambda x \lambda y.verb'xy$	(OVS)	Hixkaryana
d.	$(S \backslash NP) / NP: \lambda x \lambda y.verb'yx$	(OVS')	Päri
e.	$(S \backslash NP) \backslash NP: \lambda x \lambda y.verb'xy$	(SOV)	Turkish
f.	$(S \backslash NP) \backslash NP: \lambda x \lambda y.verb'yx$	(OSV)	Dyirbal
g.	$(S / NP) / NP: \lambda x \lambda y.verb'xy$	(VOS)	Tagalog
h.	$(S / NP) / NP: \lambda x \lambda y.verb'yx$	(VSO)	Welsh

A Micro Word-order Thought Experiment

- Syntax is a HIDDEN variable for the child.
- She has to infer a CATEGORY from form-meaning pairs only.
- These pairs are subject to universal combinatorics. (PARSING)
- These inferences set up a MENTAL GRAMMAR.
- Such grammars are necessarily PROBABILISTIC.
- All word orders are available to the child IN THE BEGINNING.

<http://tinyurl.com/CCGlab-tool>

L: Logical Form

S: Sentence

D: Derivation

- Hidden variable problem: θ are weights of lexical assumptions

$$\arg \max_L P(L \mid S; \bar{\theta}) = \arg \max_L \sum_D P(L, D \mid S; \bar{\theta}) \quad (1)$$

- Ds are inferred, not observed.
- Relating probabilities and weights (hence categories):

$$P(L, D \mid S; \bar{\theta}) = \frac{e^{\bar{f}(L,D,S) \cdot \bar{\theta}}}{\sum_L \sum_D e^{\bar{f}(L,D,S) \cdot \bar{\theta}}} \quad (2)$$

Parameter estimation: weight update

- Assume a mental grammar.
- Assume some data to be true (training set)
- What is assumed is form-meaning correspondence, not syntactic knowledge.
- Parse-to-learn to re-estimate the parameters

Learning workflow

TRAINING SET

threw the dog mommy : throw (def dog) mommy;
the dog ate the biscuit : eat (def biscuit) (def dog);
the cat ate the biscuit : eat (def biscuit) (def cat);
mommy ate the biscuit : eat (def biscuit) mommy;
the biscuit ate mommy : eat (def biscuit) mommy; % not favor OVS
mommy threw the ball : throw (def ball) mommy;
mommy threw the dog : throw (def dog) mommy;

0.1 ate v := (s\np)\np: \x\y.eat x y; % SVO
0.1 ate v := (s\np)\np: \x\y.eat y x; % SVO'
0.1 ate v := (s\np)\np: \x\y.eat y x; % OVS
0.1 ate v := (s\np)\np: \x\y.eat x y; % OVS'
0.1 ate v := (s\np)\np: \x\y.eat x y; % SOV
0.1 ate v := (s\np)\np: \x\y.eat y x; % OSV
0.1 ate v := (s\np)\np: \x\y.eat x y; % VOS
0.1 ate v := (s\np)\np: \x\y.eat y x; % VSO

GRAMMAR

PARSER

2.46 ate v := (s\np)\np: \x\y.eat x y; % SVO
2.71 ate v := (s\np)\np: \x\y.eat y x; % SVO'
-0.07 ate v := (s\np)\np: \x\y.eat y x; % OVS
-0.23 ate v := (s\np)\np: \x\y.eat x y; % OVS'
-1.04 ate v := (s\np)\np: \x\y.eat x y; % SOV
-1.04 ate v := (s\np)\np: \x\y.eat y x; % OSV
-1.04 ate v := (s\np)\np: \x\y.eat x y; % VOS
-1.04 ate v := (s\np)\np: \x\y.eat y x; % VSO

updated GRAMMAR

Log-linear model of structure prediction

$$O(\bar{\theta}) = \sum_{i=1}^n \log P(L_i \mid S_i; \bar{\theta}) = \sum_{i=1}^n \log \left(\sum_D P(L_i, D \mid S_i; \bar{\theta}) \right) \quad (3)$$

Derivative: How well each parameter serves training

$$\frac{\partial O}{\partial \theta_j} = E_{f_j(L_i, D, S_i)P(D|S_i, L_i; \bar{\theta})} - E_{f_j(L, D, S_i)P(L, D|S_i; \bar{\theta})} \quad (4)$$

Which is

$$\frac{\partial O}{\partial \theta_j} = \sum_{i=1}^n \sum_D f_j(L_i, D, S_i)P(D | S_i, L_i; \bar{\theta}) - \sum_{i=1}^n \sum_L \sum_D f_j(L, D, S_i)P(L, D | S_i; \bar{\theta}) \quad (5)$$

Then we apply Stochastic Gradient to re-estimate

Initialize $\bar{\theta}$ to some value. (6)

for $k = 0 \dots N - 1$

for $i = 1 \dots n$

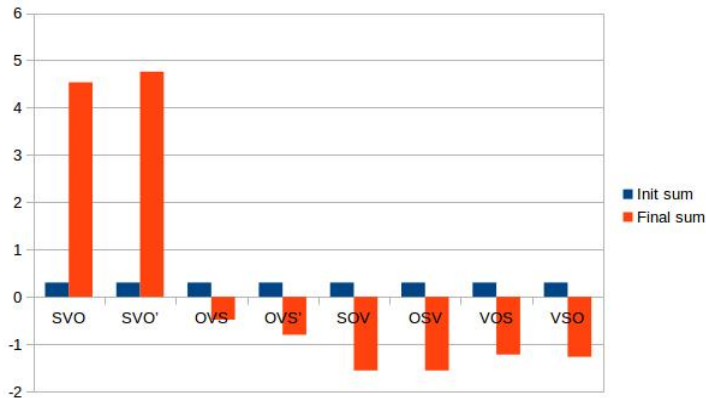
$$\bar{\theta} = \bar{\theta} + \frac{\alpha_0}{1 + c(i + kn)} \frac{\partial \log P(L_i | S_i; \bar{\theta})}{\partial \bar{\theta}}$$

- All possibilities for all verbs are entered in initial mental grammar.
- Give a set of utterances whose meanings are assumed to be known.
- Train the categories on probabilities coming from the parses.
- Some will increase, some decrease, some stay the same.
- How big is the difference from initial grammar?

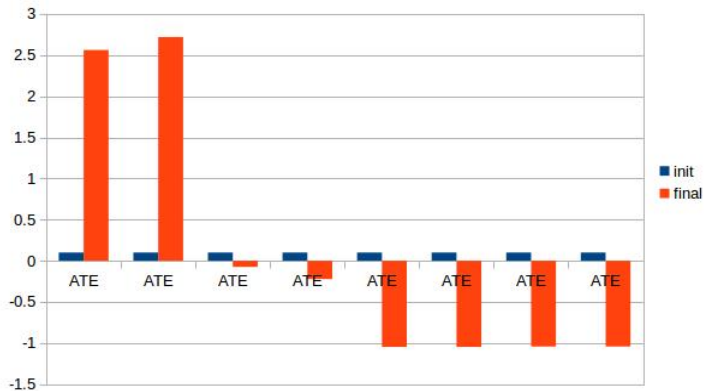
The training set

```
the dog ate the biscuit : eat (def biscuit) (def dog);
the cat ate the biscuit : eat (def biscuit) (def cat);
mommy ate the biscuit : eat (def biscuit) mommy;
the biscuit ate mommy : eat (def biscuit) mommy;
mommy threw the ball : throw (def ball) mommy;
mommy threw the dog : throw (def dog) mommy;
threw the dog mommy : throw (def dog) mommy;
the dog saw mommy : see mommy (def dog) ;
mommy saw the dog : see (def dog) mommy;
the dog walked : walk (def dog);
the cat slept : sleep (def cat);
the cat walked : walk (def cat);
the dog : def dog;
the cat : def cat;
mommy : mommy;
```

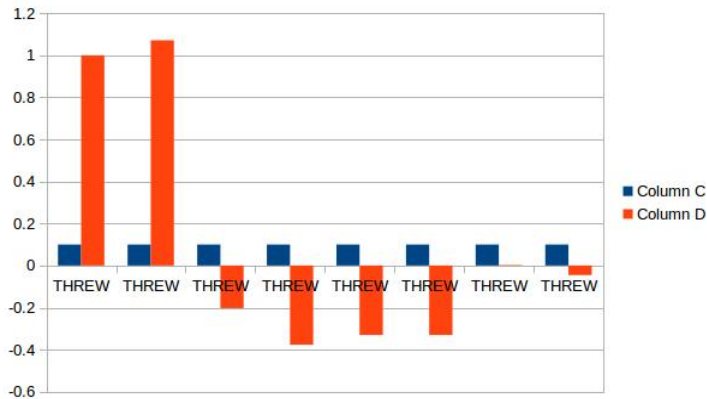

Verb categories update: 3 verbs total \times 8 categories



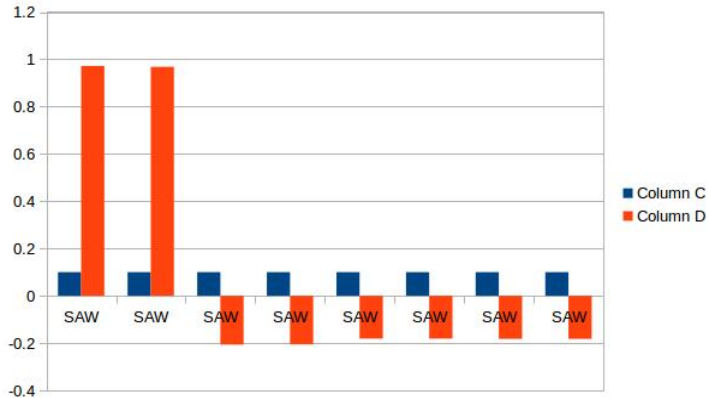
Ate parameters



Threw parameters



Saw parameters



- These conditions are quite realistic for the child.
- A computationally efficient grammatical theory can attack the hidden-variable problem in syntax
 - to fast convergence without brain switches,
 - to meet the timeframe of language acquisition and the size of grammars children learn.

Ta!