Incrementality in Compositional Distributional Semantics

M. Sadrzadeh, EECS, QMUL

SemDial 2018
joint work with M. Purver, J. Hough, R. Kempson

SYCO2, Glasgow
December 2018

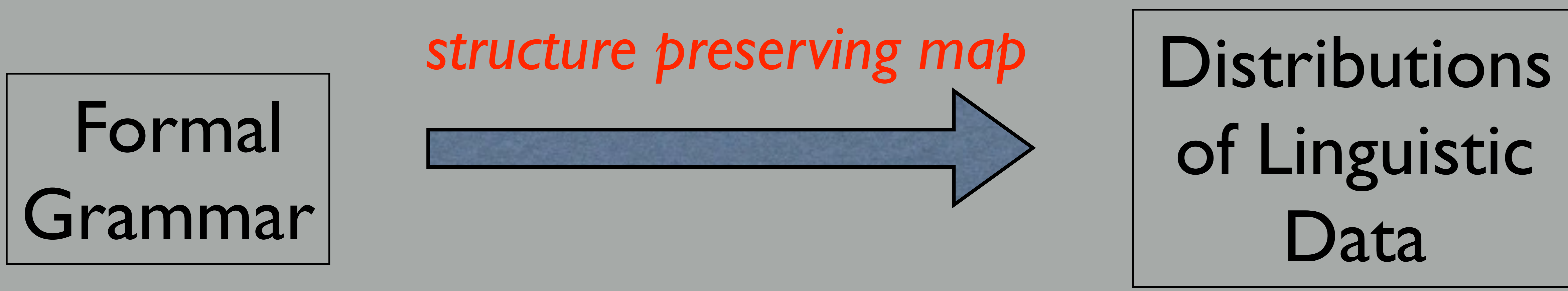
NLP in one slide

Formal Grammar $\xrightarrow{\textit{structure preserving map}}$ Semantic Calculus

NLP in one slide

Formal Grammar $\xrightarrow{\textit{structure preserving map}}$ Models of First Order Logic

NLP in one slide



Distributional Semantics

sugar, a sliced lemon, a tablespoonful of	**apricot**	preserve or jam, a pinch each of,
their enjoyment. Cautiously she sampled her first	**pineapple**	and another fruit whose taste she likened
well suited to programming on the digital	**computer**.	In finding the optimal R-stage policy from
for the purpose of gathering data and	**information**	necessary for the study authorized in the

	computer	**data**	**pinch**	**result**	**sugar**
apricot	0	0	2.25	0	2.25
pineapple	0	0	2.25	0	2.25
digital	1.66	0	0	0	0
information	0	0.57	0	0.47	0

$$\mathrm{PPMI}(w,c) = \max(\log_2 \frac{P(w,c)}{P(w)P(c)}, 0)$$

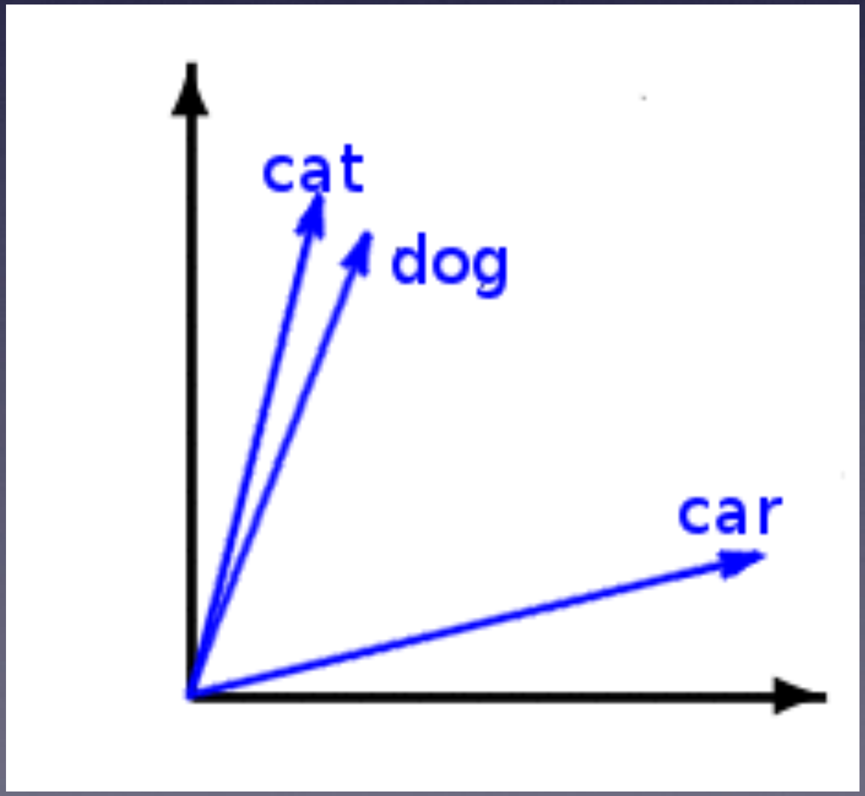
Speech and Language Processing,
Jurafsky and Martin

State of the art NLP packages

spaCy

```
import spacy
nlp = spacy.load('en_core_web_md')
tokens = nlp(u'dog cat car')
for token1 in tokens:
    for token2 in tokens:
        print(token1.text, token2.text, token1.similarity(token2))
```

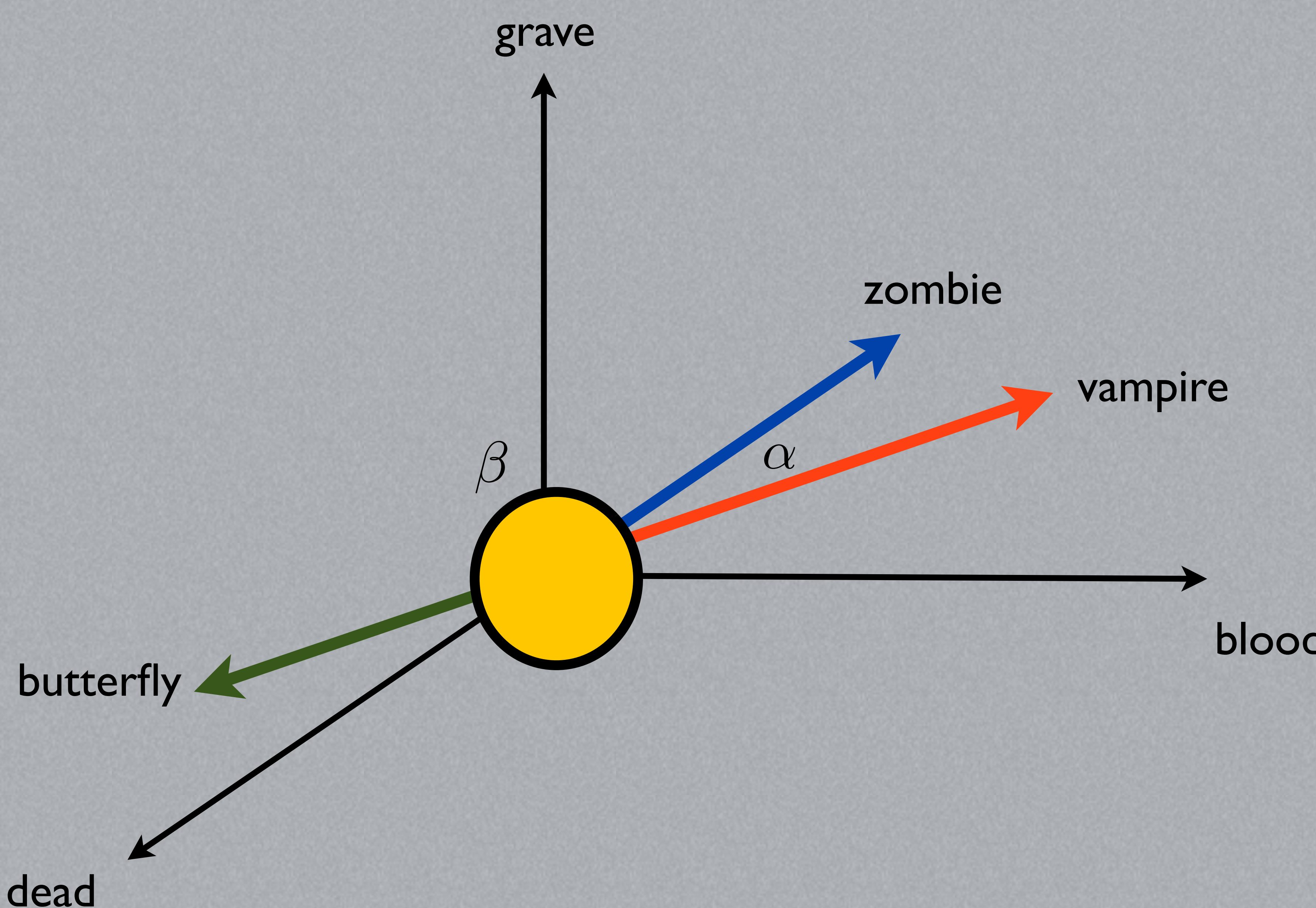
```
dog dog 1.0
dog cat 0.80168545
dog car 0.35629162
cat dog 0.80168545
cat cat 1.0
cat car 0.31907532
car dog 0.35629162
car cat 0.31907532
car car 1.0
```



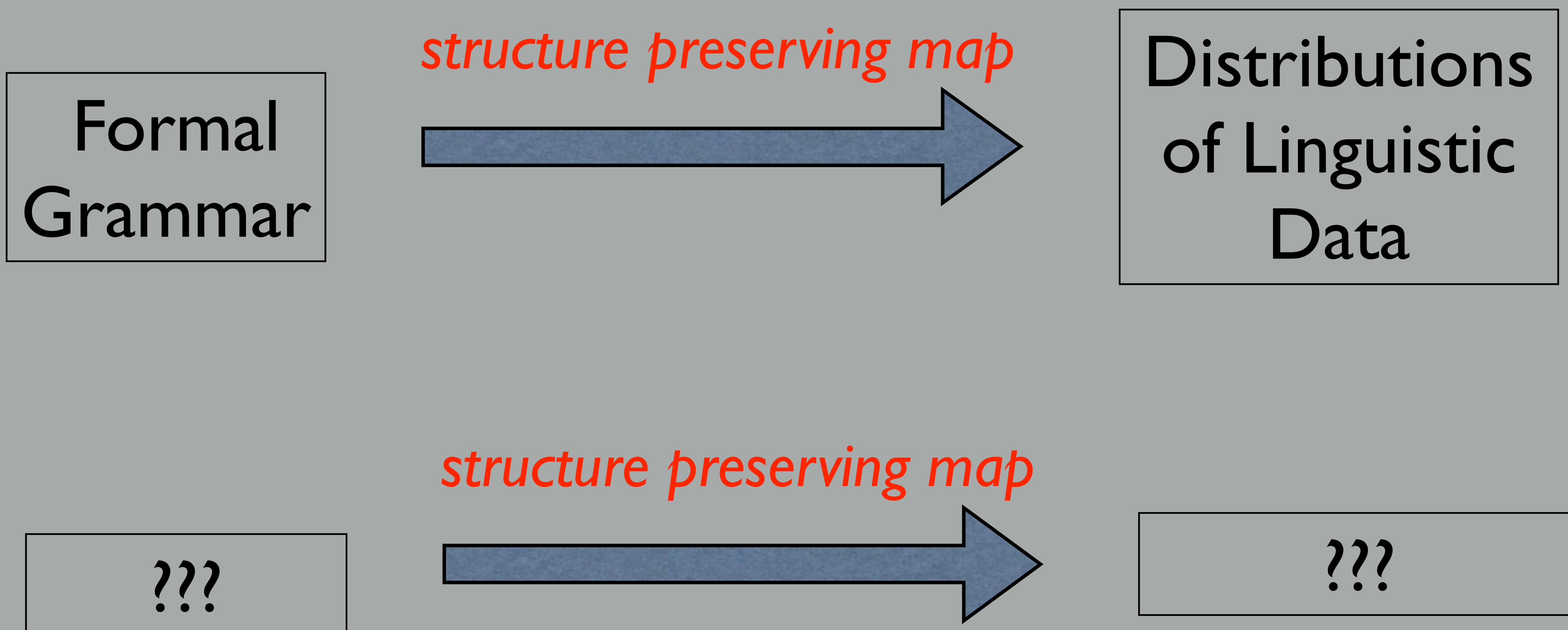
Distributional Semantics

	dog	cat	car
dog	1	0.80	0.35
cat		1	0.31
car			1

Distributional Semantics



NLP in one slide



NLP in one slide

Formal Grammar $\xrightarrow{\textit{structure preserving map}}$ Distributions of Linguistic Data

Type Grammars $\xrightarrow{\textit{structure preserving map}}$ Multilinear Algebra

CCG

Types

$$A$$

$$X/Y$$

$$X \setminus Y$$

NP, S
NP/NP, S\NP
(S\NP)/NP

noun phrase
adj, iTv
Tv

Rules

$$X/Y \quad Y \implies X$$

$$Y \quad X \setminus Y \implies X$$

NP/NP NP => NP
NP S\NP => S

Multilinear Algebraic Semantics

Vectors

$$A \mapsto \mathcal{A}$$

$$\mathcal{A} = \{e_i\}_i \ni T_i = \sum_i C_i e_i$$

Multilinear Algebraic Semantics

Matrices

$$A/B \mapsto \mathcal{A} \otimes \mathcal{B}$$

$$\mathcal{A} = \{e_i\}_i \quad \mathcal{B} = \{e_j\}_j$$

$$\mathcal{A} \otimes \mathcal{B} \ni T_{ij} = \sum_{ij} C_{ij}\, e_i \otimes e_j$$

Multilinear Algebraic Semantics

Cubes

$$A/(B/C) \mapsto \mathcal{A} \otimes (\mathcal{B} \otimes \mathcal{C}) \qquad \mathcal{A} = \{e_i\}_i \quad \mathcal{B} = \{e_j\}_j \quad \mathcal{C} = \{e_k\}_k$$

$$\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C} \ni T_{ijk} = \sum_{ijk} C_{ijk}\, e_i \otimes e_j \otimes e_k$$

Multilinear Algebraic Semantics

Higher order tensors

$$\mathcal{A} \otimes \mathcal{B} \otimes \cdots \otimes \mathcal{Z} \ni T_{ij\cdots w} = \sum_{ij\cdots w} C_{ij\cdots w}\, e_i \otimes e_j \otimes \cdots \otimes e_w$$

Multilinear Algebraic Semantics

Matrix Multiplication

$$A/B \quad B \implies A \quad \mapsto \quad (\mathcal{A} \otimes \mathcal{B}) \quad \mathcal{B} \implies \mathcal{A}$$

$$T_{ij} \quad T_j \xRightarrow{\text{tensor contract}} T_i$$

$$(\sum_{ij} C_{ij} e_i \otimes e_j)(\sum_i C_j e_j) = \sum_i C_{ij} C_j e_i \langle e_j \mid e_j \rangle$$

Multilinear Algebraic Semantics

Higher order tensor contraction

$$\cdots \mapsto \mathcal{A} \otimes \mathcal{B} \otimes \cdots \otimes \mathcal{M} \quad \mathcal{M} \otimes \mathcal{N} \otimes \mathcal{P} \otimes \cdots \otimes \mathcal{W}$$

$$T_{ij\cdots m} \quad T_{mnp\cdots w} \overset{\text{tensor contract}}{\Longrightarrow} T_{ij\cdots np\cdots w}$$

$$\begin{aligned}
&\Big(\sum_{ij\cdots m} C_{ij\cdots m} e_i \otimes e_j \otimes \cdots \otimes e_m\Big)\Big(\sum_{mn\cdots w} C_{mn\cdots w} e_m \otimes e_n \otimes \cdots \otimes e_w\Big) \\
&\qquad = \sum_{ij\cdots n\cdots w} C_{ij\cdots m} C_{mn\cdots w} e_i \otimes e_j \otimes \cdots \otimes e_n \otimes \cdots \otimes e_w \langle e_m \mid e_m \rangle
\end{aligned}$$

$$
\begin{array}{cccc}
\text{Dogs} & \text{Chase} & \text{White} & \text{Cats} \\
NP & (S \setminus NP)/NP & NP/NP & NP \\
\cline{3-4}
 & & \multicolumn{2}{c}{NP} \\
\cline{2-4}
 & \multicolumn{3}{c}{S \setminus NP} \\
\hline
\multicolumn{4}{c}{S}
\end{array}
$$

Dogs Chase White Cats

$$
\dfrac{\mathcal{N} \quad \dfrac{(\mathcal{S} \otimes \mathcal{N}) \otimes \mathcal{N} \quad \dfrac{\mathcal{N} \otimes \mathcal{N} \quad \mathcal{N}}{\mathcal{N}}}{\mathcal{S} \otimes \mathcal{N}}}{\mathcal{S}}
$$

Dogs Chase White Cats

$$\dfrac{T_i \qquad \dfrac{T_{ijk} \qquad \dfrac{T_{kl} \qquad T_l}{T_k}}{T_{ij}}}{T_j}$$

Pregroup Grammars

Types ...

$$XY^l \qquad NPNP^l \qquad NP^rS$$

$$Y^rX \qquad NP^rSNP^l$$

Rules

$$XY^lY \leq X \qquad NPNP^lNP \leq NP$$

$$YY^rX \leq X \qquad NPNP^rS \leq S$$

Catgorical Semantics

Formal Grammar $\xrightarrow{\textit{structure preserving map}}$ Distributions of Linguistic Data

Pregroup Grammars $\xrightarrow{\textit{monoidal functor}}$ FVect

Categorial Grammars
+
Distributional Semantics

Coecke, Sadrzadeh, Clark, 2010
Grefenstette and Sadrzadeh 2011, 2015
Maillard, Clark, Grefenstette, 2014
Krishnamurti and Mitchell, 2014
Baroni and Zamparelli 2010
Wijnholds (and Moortgat) 2015-16

Language Processing

Complete Sentences

Naturally Occurring Dialogue

- *“A: mary likes ...” “B: john?”*
- *“A: mary likes ...” “B: who?”*
- *“A: mary likes ...” “B: nobody really”*

Naturally Occurring Dialogue

A: Ray destroyed . . .

B: . . . the fuchsia. He never liked it. The roses he spared . . .

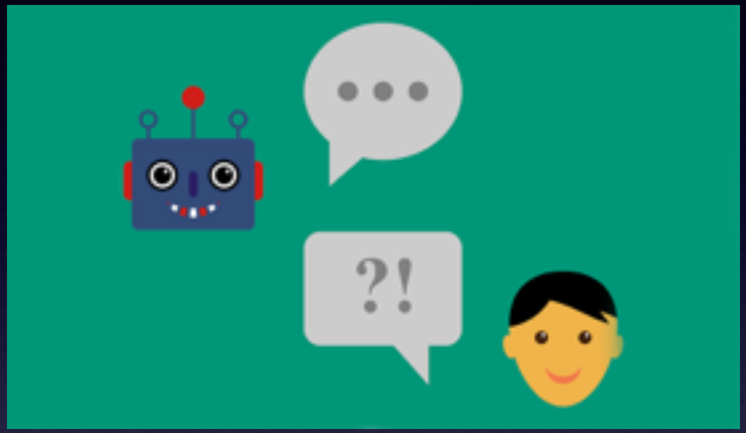
A: . . . this time.

Naturally Occurring Dialogue

A: You are going to write the letter?
B: Only if you post it!

Howes et al, 2011, Poesio and Reiser 2010

Computational Dialogue Systems



A: I want to book a ticket …
B: … from where?
A: London
B: … to where?
A: to Paris.

Purver and Kempson 2011
Purver, Eshghi, Hough 2017

Psycholinguistic Analysis

A:The footballer dribbled …
B (thinking) it means controlling the ball
A: … the ball across the pitch

A:The baby dribbled … the milk all over the floor.

Pickering and Frisson 2001

Cognitive Neuroscience

Predictive Processing: agents incrementally generate expectations and judge the degree to which they are met.

Frisson and Frith 2001
Clarke 2015

- Incremental Language Processing

Dynamic Syntax
+
Type Theoretic Semantics

Ruth Kempson, Wilfried Meyer-Viol, and Dov Gabbay. 2001.

Hough 2015, Purver et al 2014.

Recent Contribution

Dynamic Syntax
+
Distributional Semantics

Sadrzadeh, Purver, Hough, Kempson

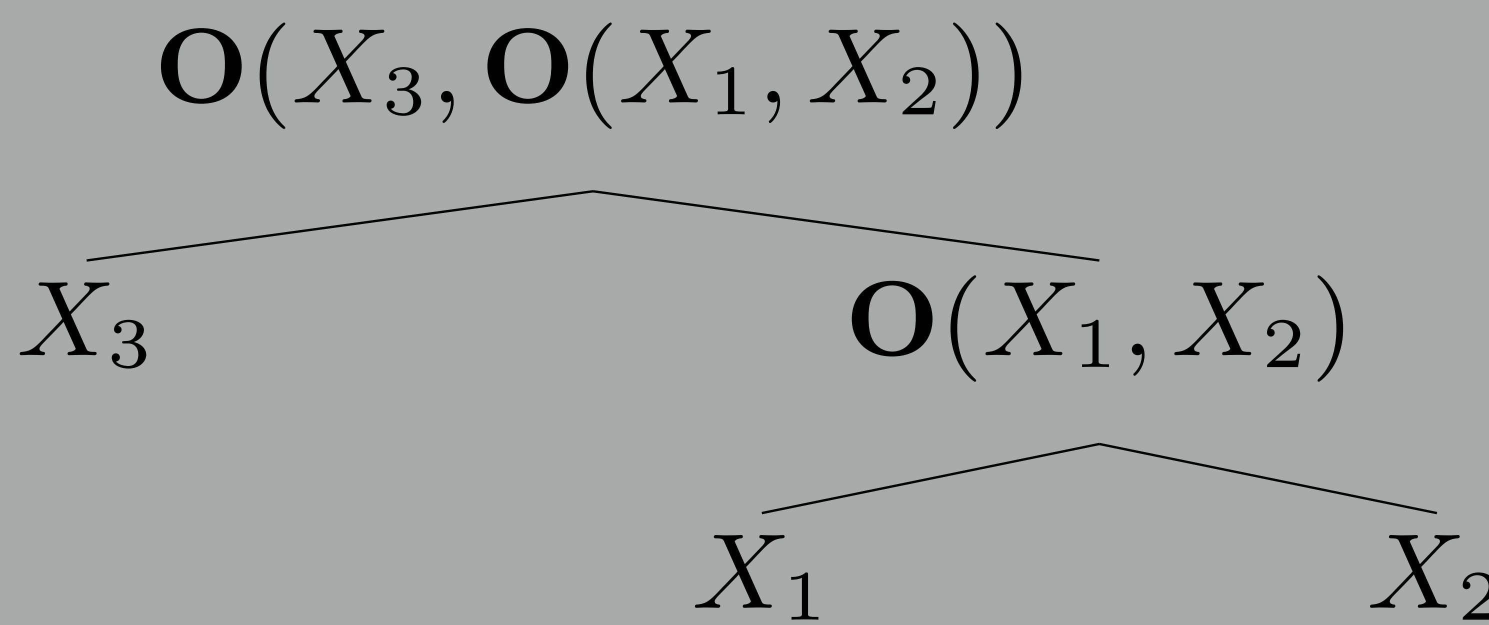
SemDial 2018

Outline

- Dynamic Syntax: DS
- CDS for DS
- Some Examples
- Some Experimental Results

Dynamic Syntax

Trees decorated with semantic formulae and applications



Dynamic Syntax

and with …

- Ty: types of formulae
- ?: requirements for further development
- <>: node currently under development
- links: connect trees of arguments of conjunctives etc

Dynamic Syntax

"mary ..."

$$?Ty(t)$$

$$Ty(e), Fo(mary) \qquad ?Ty(\langle e, t\rangle), \Diamond$$

Dynamic Syntax

"mary likes …"

$$?Ty(t)$$

$$Ty(e), Fo(mary) \qquad ?Ty(\langle e, t\rangle)$$

$$?Ty(e), \Diamond \qquad Ty(\langle e, \langle e, t\rangle\rangle), Fo(\lambda y \lambda x.like(x, y))$$

Dynamic Syntax

"mary likes john"

$$?Ty(t)$$

$$Ty(e), Fo(mary) \qquad ?Ty(\langle e, t\rangle), \Diamond$$

$$Ty(e), Fo(john) \qquad Ty(\langle e, \langle e, t\rangle\rangle), Fo(\lambda y \lambda x.like(x, y))$$

Dynamic Syntax

"mary likes john"

$$Ty(t), Fo(like(mary, john)), \Diamond$$

$$Ty(e), Fo(mary) \qquad Ty(\langle e, t\rangle), Fo(\lambda x.like(x, john))$$

$$Ty(e), Fo(john) \qquad Ty(\langle e, \langle e, t\rangle\rangle), Fo(\lambda y \lambda x.like(x, y))$$

Mary who sleeps snores.

"mary, … "

$$?S$$

$$W, T_i^{mary}, \Diamond \qquad ?W \otimes S$$

"… who … "

$$?S$$

$$W, T_i^{mary} \qquad ?W \otimes S$$

$$?S$$

$$W, T_i^{mary}, \Diamond$$

"… sleeps, … "

$$?S$$

$$W, T_i^{mary} \qquad ?W \otimes S, \Diamond$$

$$S, T_i^{mary} T_{ij}^{sleep}$$

$$W, T_i^{mary} \qquad W \otimes S, T_{ij}^{sleep}$$

"… snores … "

$$S, \mu(T_i^{mary} T_{ij}^{sleep}, T_i^{mary} T_{ij}^{snore}), \Diamond$$

$$W, T_i^{mary} \qquad W \otimes S, T_{ij}^{snore}$$

$$W, T_i^{mary} T_{ij}^{sleep}$$

$$W, T_i^{mary} \qquad W \otimes S, T_{ij}^{sleep}$$

Multilinear Algebraic Semantics for DS

Simple Nodes

$$\mathbf{O}(X_3, \mathbf{O}(X_1, X_2))$$

$$X_3 \qquad \mathbf{O}(X_1, X_2)$$

$$X_1 \qquad X_2$$

$$\begin{array}{lcll}
X_1 & \mapsto & T_{i_1 i_2 \cdots i_n} & \in \quad V_1 \otimes V_2 \otimes \cdots V_n \\
X_2 & \mapsto & T_{i_n i_{n+1} \cdots i_{n+k}} & \in \quad V_n \otimes V_{n+1} \otimes \cdots V_{n+k} \\
X_3 & \mapsto & T_{i_{n+k} i_{n+k+1} \cdots i_{n+k+m}} & \in \quad V_{n+k} \otimes V_{n+k+1} \otimes \cdots V_{n+k+m}
\end{array}$$

Multilinear Algebraic Semantics for DS

Operations Nodes

$$\begin{array}{c} \mathbf{O}(X_3, \mathbf{O}(X_1, X_2)) \\ \swarrow \quad \searrow \\ X_3 \qquad \mathbf{O}(X_1, X_2) \\ \qquad\quad \swarrow \quad \searrow \\ \qquad\quad X_1 \qquad X_2 \end{array}$$

$$\begin{aligned} \mathbf{O}(X_1, X_2) \quad & \mapsto \quad T_{i_1 i_2 \cdots i_n} T_{i_n i_{n+1} \cdots i_{n+k}} \\ & \in \quad V_1 \otimes V_2 \otimes \cdots \otimes V_{n-1} \otimes V_{n+1} \otimes \cdots \otimes V_{n+k} \end{aligned}$$

Multilinear Algebraic Semantics for DS

Extras

- Ty(t) —--> S
- Ty(e) —> W
- ?X —--> sum or direct sum of the words and phrase with semantics in X and their probabilities
- —--> a neutral element such as the identity in X
- —> a tensor full of 1's in X

“mary ...”

$$?S$$

$$W \ni T_i^{mary} \qquad ?W \otimes S, \Diamond$$

“mary likes ...”

$$?S$$

$$W \ni T_i^{mary} \qquad ?W \otimes S$$

$$?W, \diamondsuit \qquad \underline{W \otimes S \otimes W \ni T_{ijk}^{like}}$$

“mary likes john”

$$S \ni T_i^{mary} T_{ijk}^{like} T_k^{john}, \diamondsuit$$

$$W \ni T_i^{mary} \qquad W \otimes S \ni T_{ijk}^{like} T_k^{john}$$

$$W \ni T_k^{john} \qquad W \otimes S \otimes W \ni T_{ijk}^{like}$$

Incremental Utterances

"Babies ..."

$$?S$$

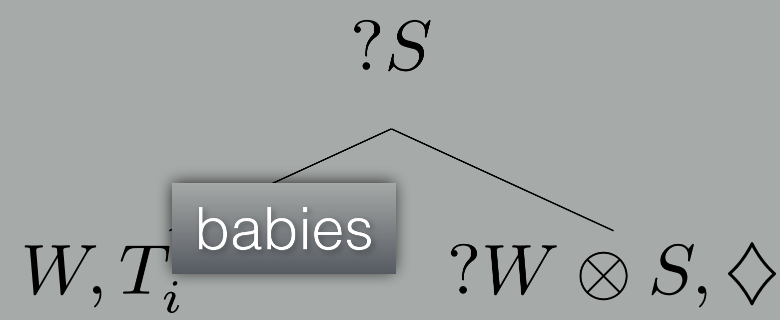
babies

$$W, T_i \qquad ?W \otimes S, \Diamond$$

$$T_i^{babies} T_{ij}^{+}$$

Incremental Utterances

"Babies ..."



$$T^{+}_{ij} =$$

Incremental Utterances

"Babies …"

$?S$

babies

W, T_i $\quad$ $?W \otimes S, \diamondsuit$

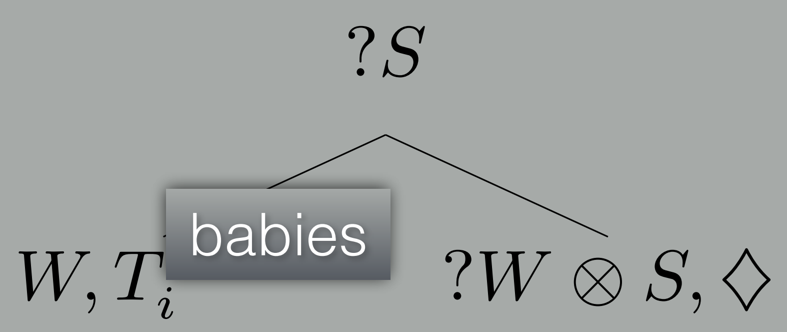
$$T_{ij}^{+} =$$

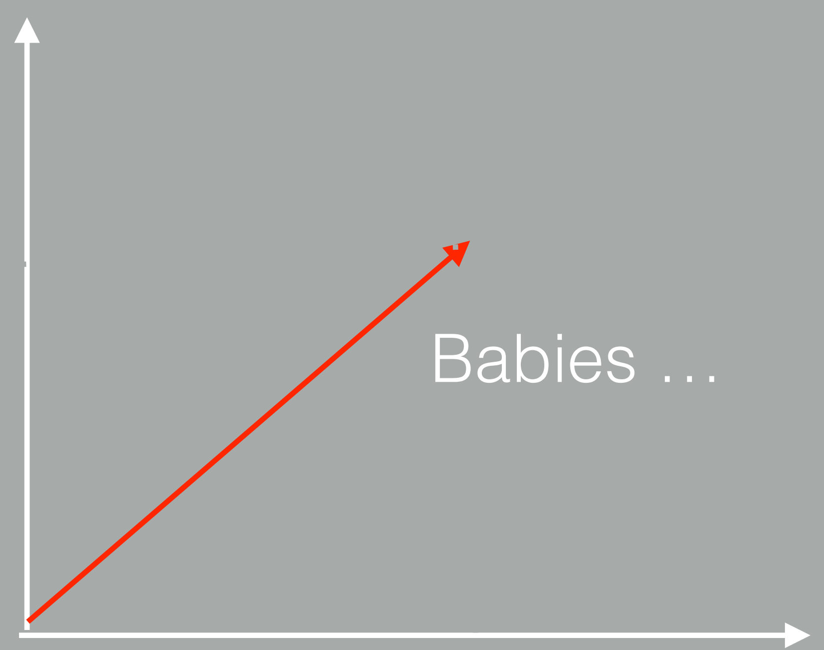
$$T^{vomit} + T^{score} + T^{dribble} + T^{control\ baby} + T^{control\ milk} + T^{control\ footballer} + T^{control\ ball}$$

Incremental Utterances

"Babies ..."

$$T_i^{babies} T_{ij}^{+}$$

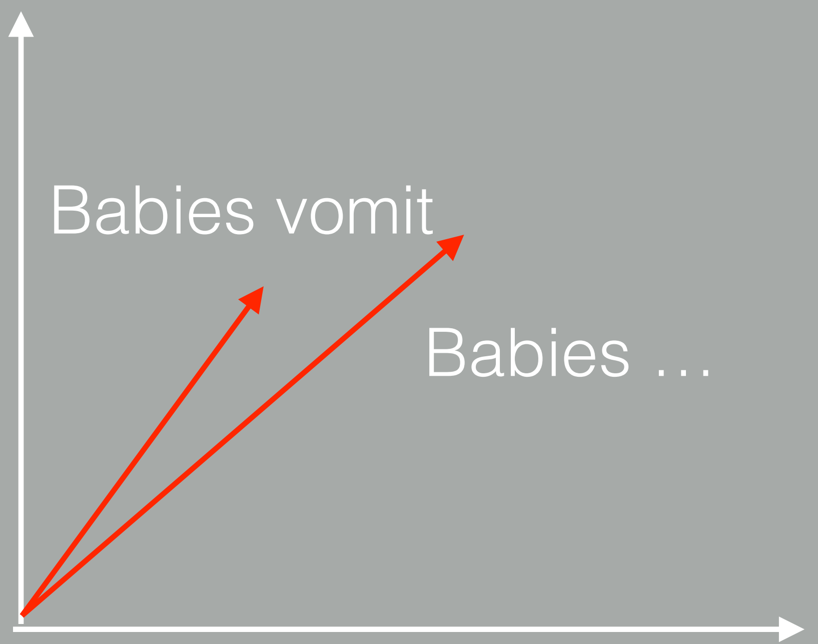




Incremental Utterances

"Babies vomit"

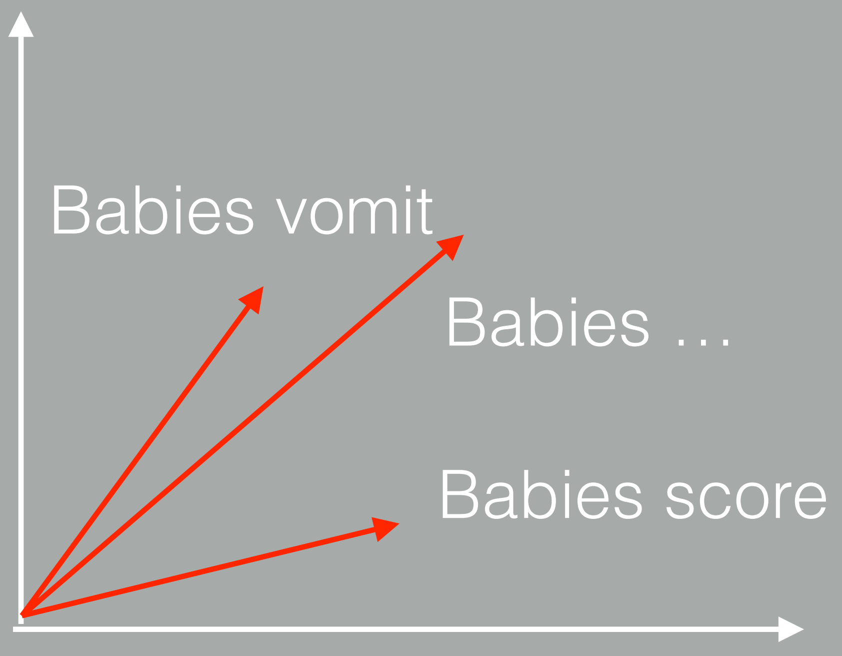
$$T_i^{babies} T_{ij}^{vomit}$$



Incremental Utterances

"Babies score"

$$T_i^{babies} T_{ij}^{score}$$



Incremental Utterances

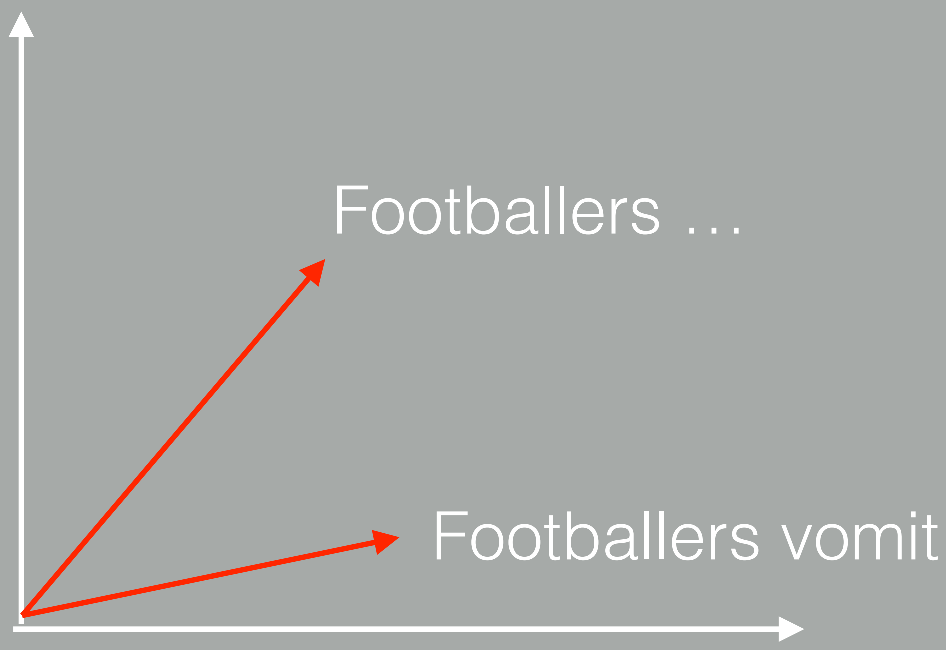
“Footballers …”

$$T^{footballers}T^{+}$$

Incremental Utterances

"Footballers vomit"

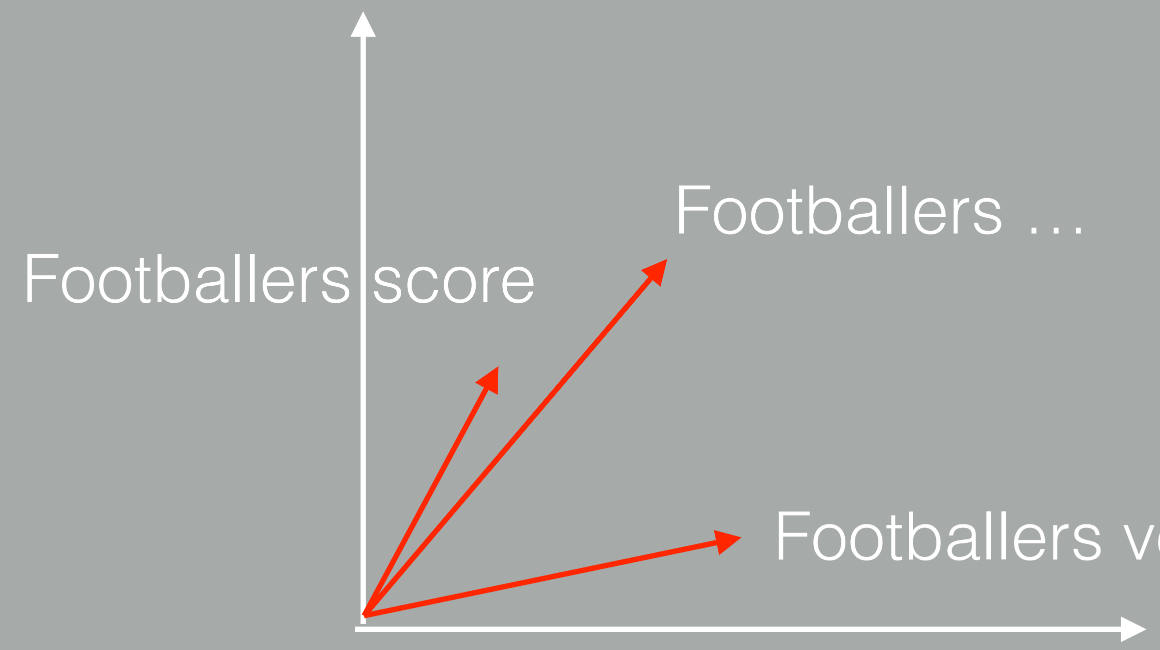
$$T^{footballers} T_{ij}^{vomit}$$



Incremental Utterances

"Footballers score"

$$T_i^{footballers} T_{ij}^{score}$$



Dataset

- Kartsaklis D., MS, Pulman S.: Separating disambiguation from composition in compositional distributional semantics.
- Chose ambiguous verbs and two landmark meanings from Pickering and Frisson 2001
- Picked subjects and objects for landmarks using most frequently occurring ones in the BNC

Dataset

- Pairs of subjects and complete sentences

  (footballers … , footballers dribble milk)
  (footballers … , footballers dribble ball)

- Pairs of subject+verb and complete sentences:

  (footballers dribble … , footballers dribble milk)
  (footballers dribble … , footballers dribble ball)

- Pairs of complete sentences:

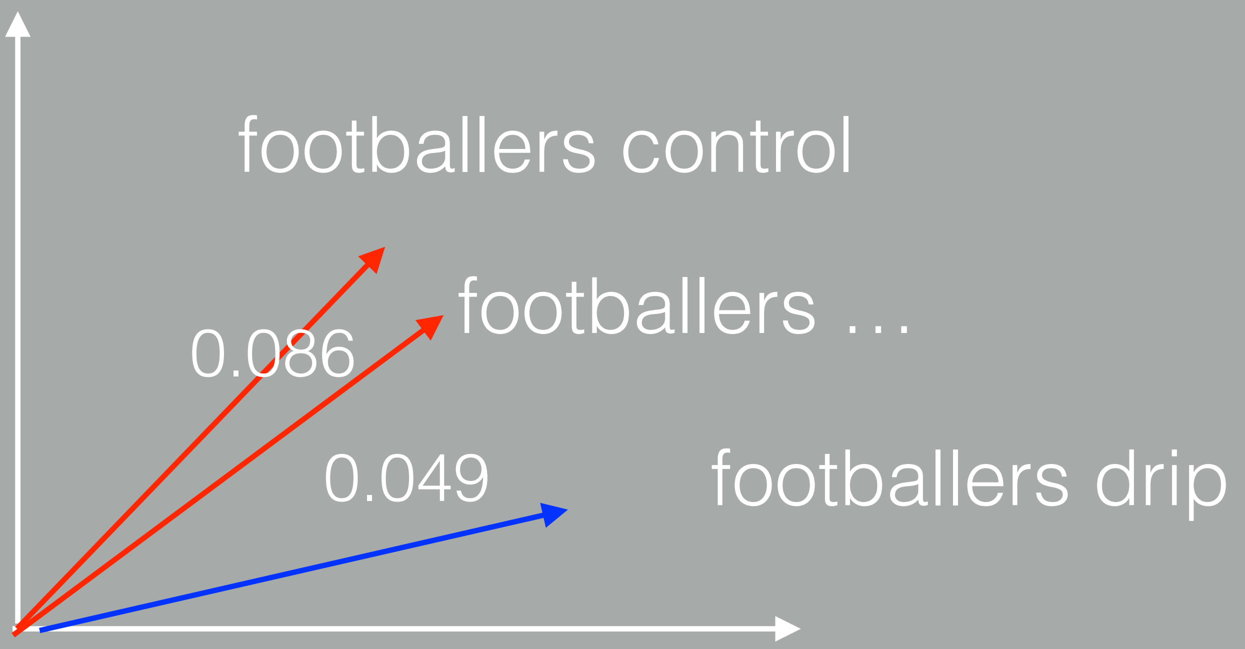
  (footballers dribble milk , footballers dribble ball)
  (babies dribble milk , babies dribble ball)

Data

Vectors: 300 Dim from Word2Vec,
Tensors: the G&S EMNLP 2011 method

Data

Just subject

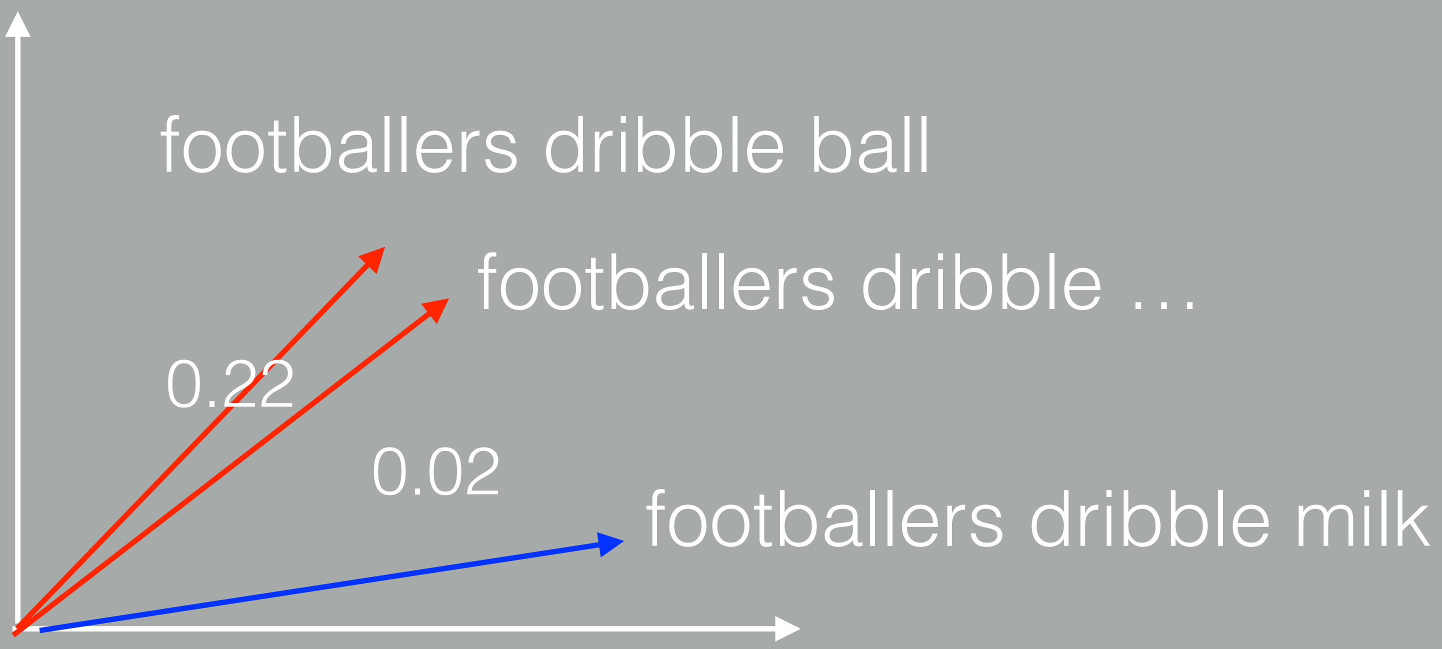


Data

Just Subject

Data

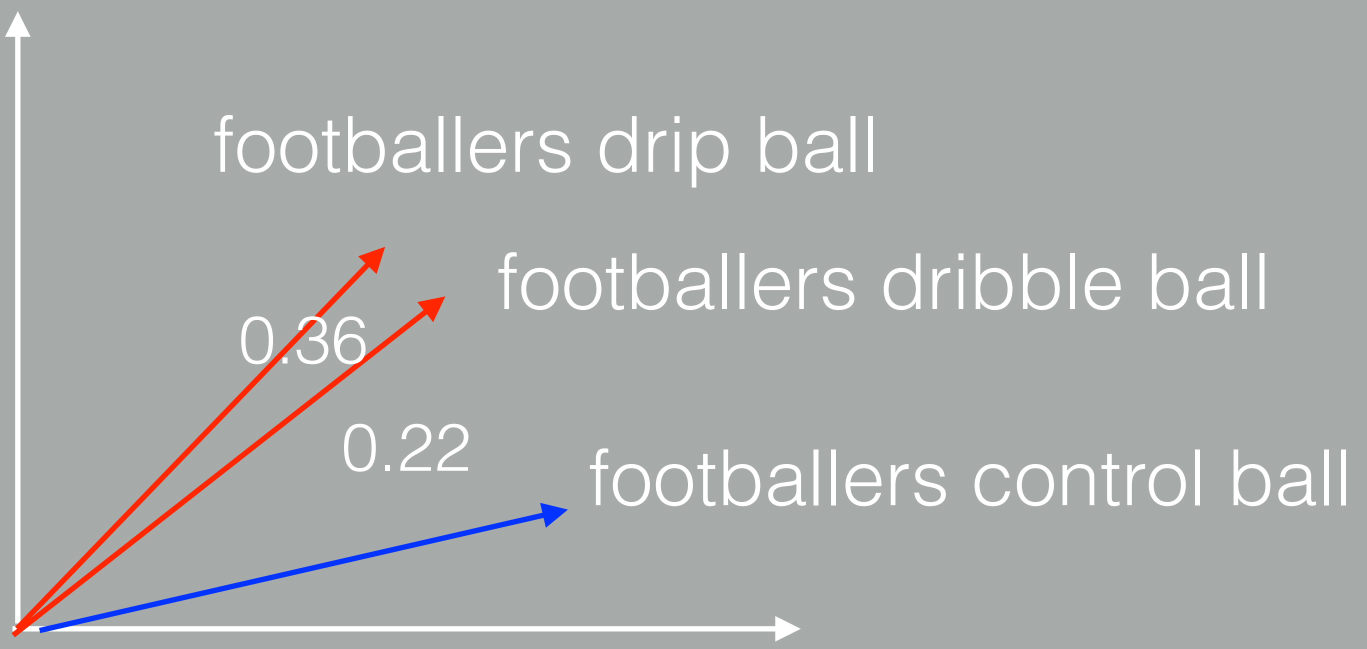
Subject + Verb



Data

Complete Sentences

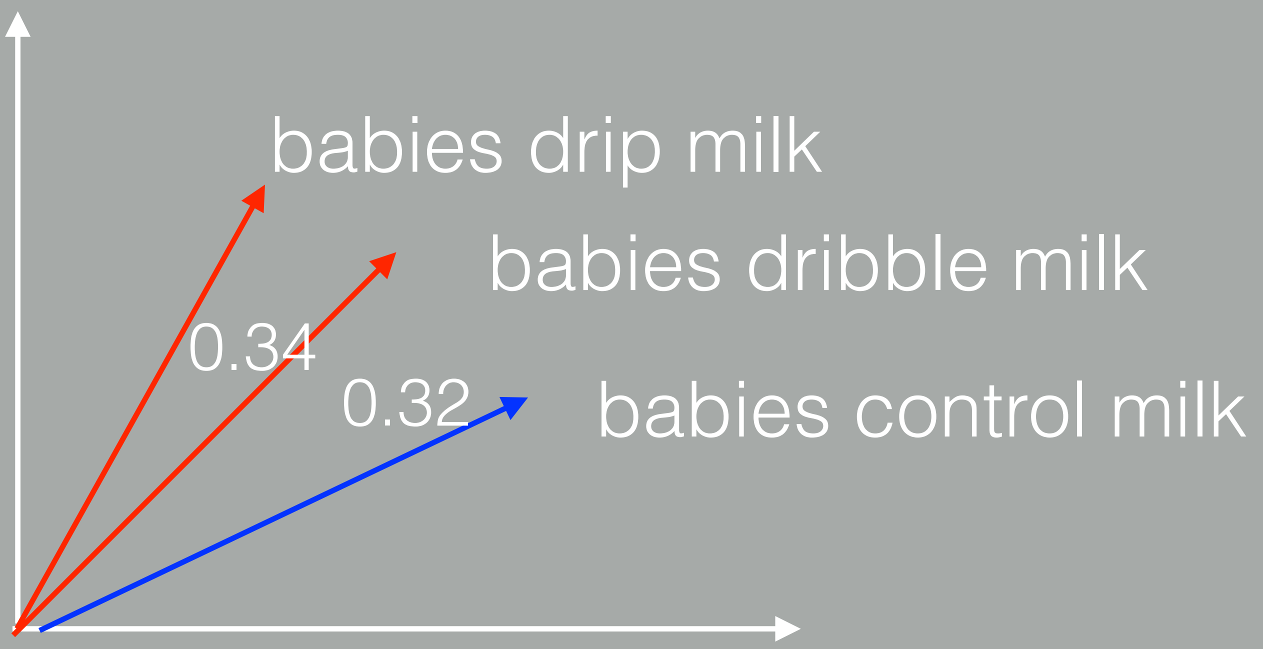
0.22 < 0.36



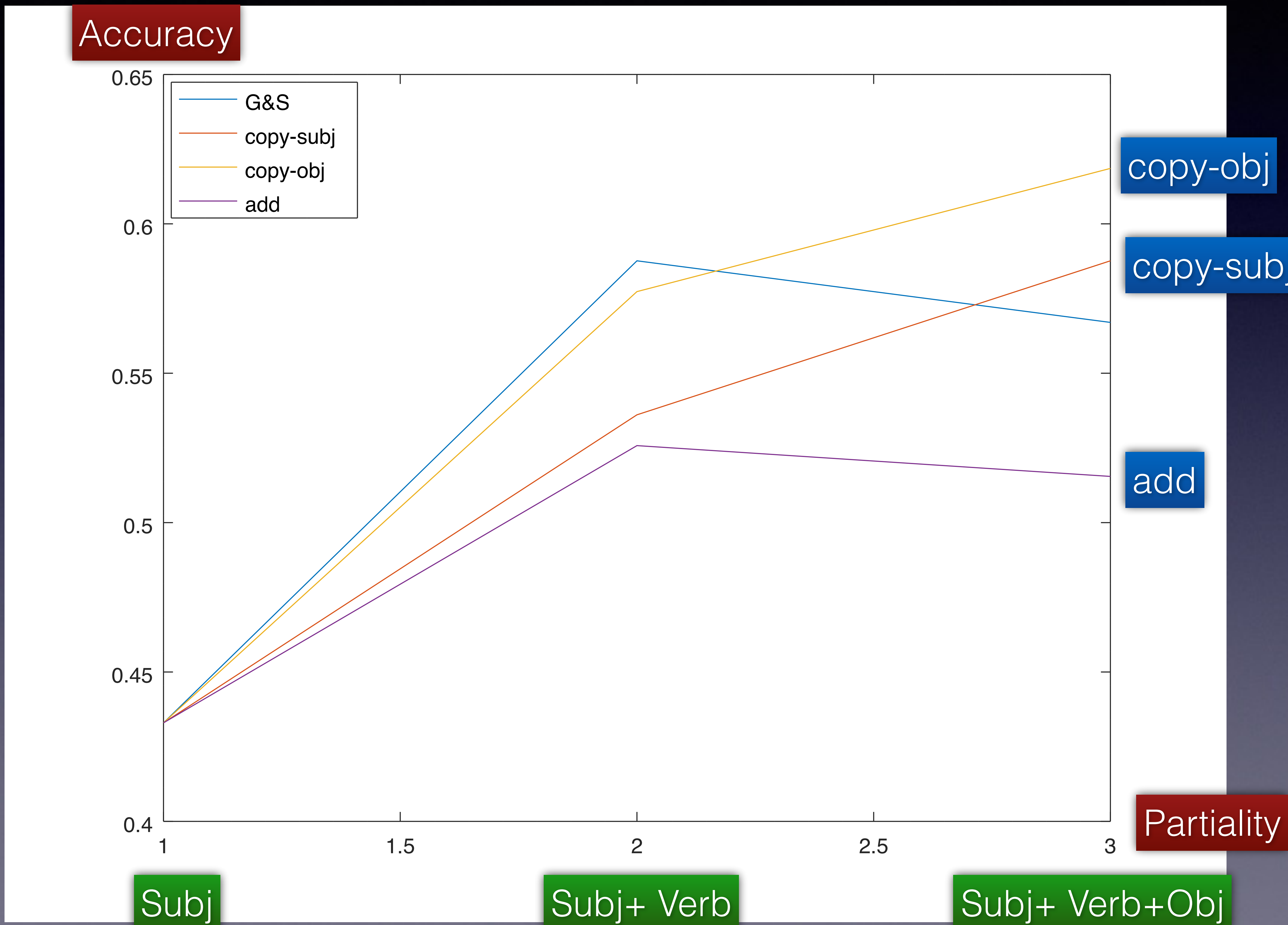
Data

Complete Sentences

0.34 > 0.32



Accuracy Results



Work in Progress

Implement the plausibilities model of Clark 2013, Polajnar et al 2015

… under way …

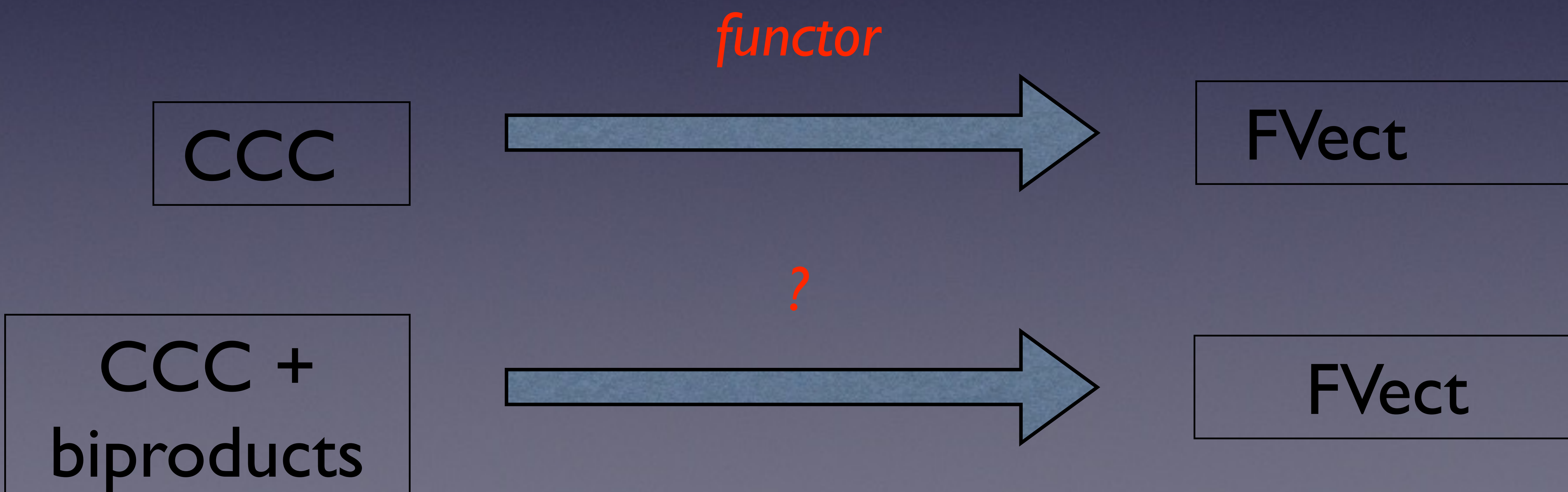
Extend it to experimental expectation predication

…

Incremental Understanding of Dialogue Content

. . .

Categorical Semantics



A: Thank …

$\otimes$

B: … you!