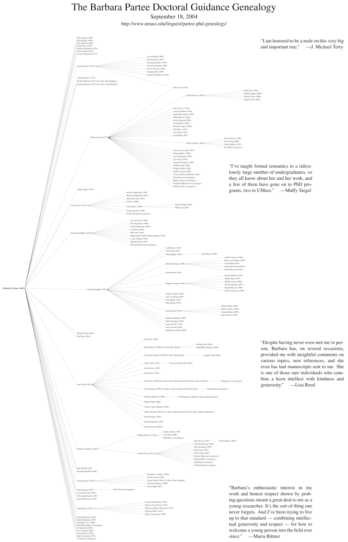
Compositionality or systematicity?

Christopher Potts

Stanford Linguistics

Franklin Institute Symposium:
The Past, Present, and Future of Formal Semantics
April 19, 2021

The thrill of scientific inquiry



Central questions

My central questions

When linguists seek compositional analyses of linguistic phenomena:

- What principles guide their investigations?
- What higher-level goals are they actually pursuing?

Where the questions lead me

- *Learning* semantic representations can lead to richer theories of language and language use,
- but compositionality is too constraining in these situations. *Systematicity* might be a better goal.

Plan

1. The compositionality principle
2. The compositionality heuristic
3. Semantic parsing
4. Recursive deep learning models
5. Conclusions

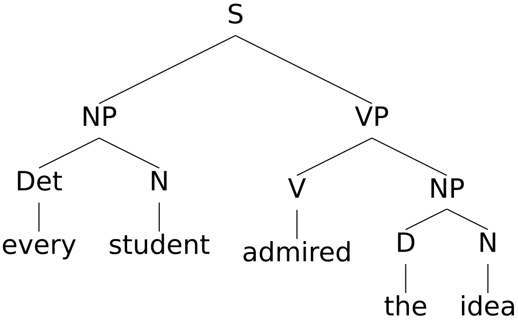
The compositionality principle

1. The compositionality principle
2. The compositionality heuristic
3. Semantic parsing
4. Recursive deep learning models
5. Conclusions

Informal statement

Compositionality

The meaning of a phrase is a function of the meanings of its immediate syntactic constituents and the way they are combined.

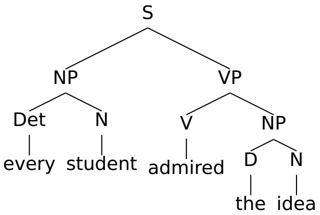


The usual motivation

1. Modeling all meaningful units

   $[\![every]\!] = \lambda f\ \lambda g\ \forall x\ ((f\ x) \rightarrow (g\ x))$

2. “Infinite” capacity
3. Creativity
4. Systematicity



Montague: Unconstrained compositionality

Partee (1996) on Montague (1970):

The central idea is that anything that should count as a grammar should be able to be cast in the following form: the syntax is an algebra, the semantics is an algebra, and there is a homomorphism mapping elements of the syntactic algebra onto elements of the semantic algebra.

[. . .]

It is the homomorphism requirement, which is in effect the compositionality requirement, that provides the most important constraint on UG in Montague’s sense [. . .].

Formal results

Zadrozny (1994)

Any meaning function (map from forms to meanings) can be turned into a compositional one in the sense of the homomorphism requirement.

Kazmi & Pelletier (1998) respond “Wait, what?”

Here is a non-compositional semantics:

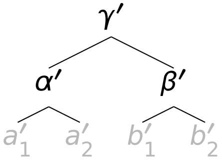
- $[\![\textit{sleep}]\!] = [\![\textit{doze}]\!]$
- $[\![\textit{sleep tight}]\!] \neq [\![\textit{doze tight}]\!]$

Zadrozny showed how to create a kind of syntactic layer where compositionality holds. A similar argument is made by Dever 1999.

Dowty's context-free compositionality

Dowty (2007):

When a rule f combines $\alpha, \beta(\ldots)$ to form γ, the corresponding semantic rule g that produces the meaning γ' of γ, from α' and β', may depend only on α' "as a whole", it may not depend on any meanings from which α' was formed compositionally by earlier derivational steps (similarly for β).



The compositionality heuristic

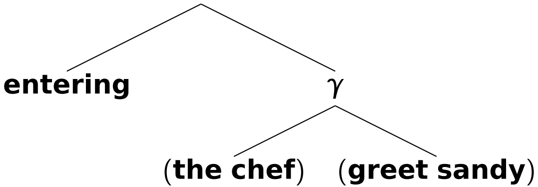
Compositionality as methodology

Janssen (1997:461)

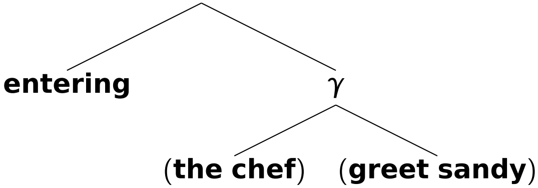
Compositionality is not a formal restriction on what can be achieved, but a methodology on how to proceed.

Example: Subjectless predicational adjuncts

Example: Subjectless predicational adjuncts

1. Entering the restaurant, the chef greeted Sandy.
2. 

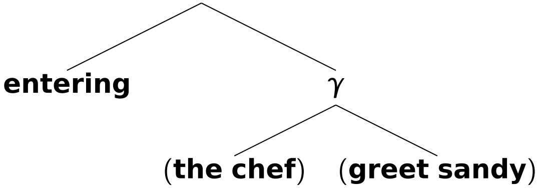
Example: Subjectless predicational adjuncts

1. Entering the restaurant, the chef greeted Sandy.
2. 

Potential rule

The implicit argument of a front subjectless predicational adjuncts must be the subject of the matrix clause.

Example: Subjectless predicational adjuncts

1. Entering the restaurant, the chef greeted Sandy.
2. 

Potential rule

The implicit argument of a front subjectless predicational adjuncts must be the subject of the matrix clause.

Assuming context-free compositionality, the rule cannot be correct as stated.

Example: Subjectless predicational adjuncts

From `http://arnoldzwicky.wordpress.com/category/danglers/`:

1. “Having been in Australia for 17 years, a foreign national wishing to work in Australia must be of good character.”
2. “Fearing a massive lay-off, there was a general sense of relief when the boss announced there would be no new budget cuts.”
3. “Rich and creamy, your guests will never guess that this pie is light.”

Example: Subjectless predicational adjuncts

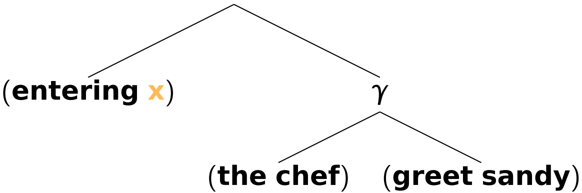
Pragmatic constraint

The free variable in a subjectless predicational adjuncts should refer to a discourse topic.

Example: Subjectless predicational adjuncts

Pragmatic constraint

The free variable in a subjectless predicational adjuncts should refer to a discourse topic.

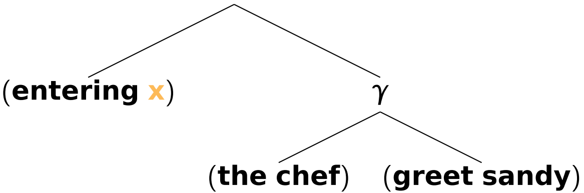
1. Entering the restaurant, the chef greeted Sandy.
2. 

Example: Subjectless predicational adjuncts

Pragmatic constraint

The free variable in a subjectless predicational adjuncts should refer to a discourse topic.

1. Entering the restaurant, the chef greeted Sandy.
2.



Subjects/topic correlation

In English, subjects are often topics.

Example: Compounds and systematicity

Partee (1995:341):

In compounds [. . .] there is no general rule for predicting the interpretation of the combination

Levin et al.'s (2019) novel compounds experiment:

Modifier	Head	Example	Event	Perceptual Environmental/
Artifact	Artifact	stew skillet	93%	7%
Natural kind	Artifact	stream wheel	88%	12%
Artifact	Natural kind	stew chickpea	66%	34%
Natural kind	Natural kind	stream vegetable	15%	85%

flat tire/beer/note/file

Semantic parsing

The semanticist's ideal

Every student attended a lecture

$$\forall z\ ((\textbf{student}\ z) \rightarrow (\exists x\ (\textbf{lecture}\ x) \wedge (\textbf{attended}\ x\ z)))$$

$$\lambda g\ \forall z\ ((\textbf{student}\ z) \rightarrow (g\ z)) \qquad \lambda Q\ (Q\ (\lambda y\ (\exists x\ (\textbf{lecture}\ x) \wedge (\textbf{attended}\ x\ y))$$

$$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z)) \quad \textbf{student}$$

$$\textbf{attended}_S \qquad \lambda g\ (\exists x\ (\textbf{lecture}\ x) \wedge (g\ x))$$

$$\lambda f\ \lambda g\ (\exists x\ (f\ x) \wedge (g\ x)) \quad \textbf{lecture}$$

The semanticist's ideal

Every student attended a lecture

$$\exists x\ (\textbf{lecture}\ x) \wedge (\forall z\ ((\textbf{student}\ z) \rightarrow (\textbf{attended}\ x\ z)))$$

$\lambda g\ \forall z\ ((\textbf{student}\ z) \rightarrow (g\ z))$ $\lambda Q\ (\exists x\ (\textbf{lecture}\ x) \wedge (Q\ (\lambda y\ (\textbf{attended}\ x\ y))))$

$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z))$ **student**

$\textbf{attended}_l$ $\lambda g\ (\exists x\ (\textbf{lecture}\ x) \wedge (g\ x))$

$\lambda f\ \lambda g\ (\exists x\ (f\ x) \wedge (g\ x))$ **lecture**

The semanticist's ideal

Every child ate a lollipop

$$\forall z\, ((\textbf{child}\ z) \rightarrow (\exists x\, (\textbf{lollipop}\ x) \wedge (\textbf{ate}\ x\ z)))$$

- $\lambda g\, \forall z\, ((\textbf{child}\ z) \rightarrow (g\ z))$
  - $\lambda f\, \lambda g\, \forall z\, ((f\ z) \rightarrow (g\ z))$
  - $\textbf{child}$
- $\lambda Q\, (Q\, (\lambda y\, (\exists x\, (\textbf{lollipop}\ x) \wedge (\textbf{ate}\ x\ y)))$
  - $\textbf{ate}_S$
  - $\lambda g\, (\exists x\, (\textbf{lollipop}\ x) \wedge (g\ x))$
    - $\lambda f\, \lambda g\, (\exists x\, (f\ x) \wedge (g\ x))$
    - $\textbf{lollipop}$

The semanticist's ideal

Every child ate a lollipop

$$\exists x\ (\textbf{lollipop}\ x) \land (\forall z\ ((\textbf{child}\ z) \rightarrow (\textbf{ate}\ x\ z)))$$

$$\lambda g\ \forall z\ ((\textbf{child}\ z) \rightarrow (g\ z)) \qquad \lambda Q\ (\exists x\ (\textbf{lollipop}\ x) \land (Q\ (\lambda y\ (\textbf{ate}\ x\ y))))$$

$$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z)) \quad \textbf{child} \qquad \textbf{ate}_l \qquad \lambda g\ (\exists x\ (\textbf{lollipop}\ x) \land (g\ x))$$

$$\lambda f\ \lambda g\ (\exists x\ (f\ x) \land (g\ x)) \quad \textbf{lollipop}$$

The semanticist's ideal

Every child ate a lollipop

$$\exists x\ (\mathbf{lollipop}\ x) \wedge (\forall z\ ((\mathbf{child}\ z) \rightarrow (\mathbf{ate}\ x\ z)))$$

$\lambda g\ \forall z\ ((\mathbf{child}\ z) \rightarrow (g\ z))$ $\quad$ $\lambda Q\ (\exists x\ (\mathbf{lollipop}\ x) \wedge (Q\ (\lambda y\ (\mathbf{ate}\ x\ y))))$

$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z))$ $\quad$ $\mathbf{child}$ $\quad$ $\mathbf{ate}_l$ $\quad$ $\lambda g\ (\exists x\ (\mathbf{lollipop}\ x) \wedge (g\ x))$

$\lambda f\ \lambda g\ (\exists x\ (f\ x) \wedge (g\ x))$ $\quad$ $\mathbf{lollipop}$

But is this really so ideal?

Crude grammars refined via learning

Crude grammars refined via learning

Chat80 (Warren & Pereira 1982):

```
/* Sentences */
sentence(S) --> declarative(S), terminator(.) .
sentence(S) --> wh_question(S), terminator(?) .
sentence(S) --> yn_question(S), terminator(?) .
sentence(S) --> imperative(S), terminator(!) .

/* Noun Phrase */
np(np(Agmt,Pronoun,[]),Agmt,NPCase,def,_,Set,Nil) -->
  {is_pp(Set)},
  pers_pron(Pronoun,Agmt,Case),
  {empty(Nil), role(Case,decl,NPCase)}.

/* Prepositional Phrase */
pp(pp(Prep,Arg),Case,Set,Mask) -->
  prep(Prep),
  {prep_case(NPCase)},
  np(Arg,_,NPCase,_,Case,Set,Mask).
```

Crude grammars refined via learning

for $w \in \textit{Words}$
    for $X \in \textit{Categories}$
        for $d \in \textit{Domain}$
            yield '$X \rightarrow w : d$'

Crude grammars refined via learning

for $w \in \mathit{Words}$
    for $X \in \mathit{Categories}$
        for $d \in \mathit{Domain}$
            yield '$X \rightarrow w : d$'

0	N	$\rightarrow$	*dog*	:	**dog**
0	V	$\rightarrow$	*dog*	:	$\mathbf{dog}_v$
0	N	$\rightarrow$	*dog*	:	**cat**
0	N	$\rightarrow$	*cat*	:	**cat**
0	N	$\rightarrow$	*cat*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**jump**

Crude grammars refined via learning

N

dog : **dog**

for $w \in Words$
    for $X \in Categories$
        for $d \in Domain$
            yield '$X \rightarrow w : d$'

1	N	→	*dog*	:	**dog**
0	V	→	*dog*	:	$\mathbf{dog}_V$
0	N	→	*dog*	:	**cat**
0	N	→	*cat*	:	**cat**
0	N	→	*cat*	:	**dog**
0	V	→	*jump*	:	**dog**
0	V	→	*jump*	:	**jump**

Crude grammars refined via learning

N

dog : **dog**

V

dog : $\mathbf{dog}_V$

```
for w ∈ Words
    for X ∈ Categories
        for d ∈ Domain
            yield 'X → w : d'
```

1	N	→	*dog*	:	**dog**
1	V	→	*dog*	:	$\mathbf{dog}_V$
0	N	→	*dog*	:	**cat**
0	N	→	*cat*	:	**cat**
0	N	→	*cat*	:	**dog**
0	V	→	*jump*	:	**dog**
0	V	→	*jump*	:	**jump**

Crude grammars refined via learning

for $w \in \textit{Words}$
  for $X \in \textit{Categories}$
    for $d \in \textit{Domain}$
      yield '$X \to w : d$'

1	N	→	*dog*	:	**dog**
1	V	→	*dog*	:	$\mathbf{dog}_V$
0	N	→	*dog*	:	**cat**
1	N	→	*cat*	:	**cat**
0	N	→	*cat*	:	**dog**
0	V	→	*jump*	:	**dog**
0	V	→	*jump*	:	**jump**

N
|
dog : **dog**

V
|
dog : $\mathbf{dog}_V$

N
|
cat : **cat**

Crude grammars refined via learning

for $w \in Words$
    for $X \in Categories$
        for $d \in Domain$
            yield ‘$X \rightarrow w : d$’

2	N	→	*dog*	:	**dog**
1	V	→	*dog*	:	**dog**$_v$
0	N	→	*dog*	:	**cat**
1	N	→	*cat*	:	**cat**
0	N	→	*cat*	:	**dog**
0	V	→	*jump*	:	**dog**
0	V	→	*jump*	:	**jump**

N
|
dog : **dog**

V
|
dog : **dog**$_v$

N
|
cat : **cat**

N
|
dog : **dog**

Crude grammars refined via learning

```
for w ∈ Words
    for X ∈ Categories
        for d ∈ Domain
            yield 'X → w : d'
```

3	N	$\rightarrow$	*dog*	:	**dog**
1	V	$\rightarrow$	*dog*	:	$\mathbf{dog}_{V}$
0	N	$\rightarrow$	*dog*	:	**cat**
1	N	$\rightarrow$	*cat*	:	**cat**
0	N	$\rightarrow$	*cat*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**jump**

N
|
dog : **dog**

N
|
dog : **dog**

V
|
dog : $\mathbf{dog}_{V}$

N
|
cat : **cat**

N
|
dog : **dog**

Crude grammars refined via learning

```
for w ∈ Words
    for X ∈ Categories
        for d ∈ Domain
            yield 'X → w : d'
```

3	N	$\rightarrow$	*dog*	:	**dog**
1	V	$\rightarrow$	*dog*	:	$\mathbf{dog}_V$
0	N	$\rightarrow$	*dog*	:	**cat**
1	N	$\rightarrow$	*cat*	:	**cat**
0	N	$\rightarrow$	*cat*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**dog**
1	V	$\rightarrow$	*jump*	:	**jump**

$$\begin{array}{c} \mathrm{N} \\ | \\ \textit{dog} : \textbf{dog} \end{array} \qquad \begin{array}{c} \mathrm{N} \\ | \\ \textit{dog} : \textbf{dog} \end{array}$$

$$\begin{array}{c} \mathrm{V} \\ | \\ \textit{dog} : \textbf{dog}_V \end{array} \qquad \begin{array}{c} \mathrm{V} \\ | \\ \textit{jump} : \textbf{jump} \end{array}$$

$$\begin{array}{c} \mathrm{N} \\ | \\ \textit{cat} : \textbf{cat} \end{array}$$

$$\begin{array}{c} \mathrm{N} \\ | \\ \textit{dog} : \textbf{dog} \end{array}$$

Crude grammars refined via learning

for $w \in \mathit{Words}$
    for $X \in \mathit{Categories}$
        for $d \in \mathit{Domain}$
            yield ‘$X \rightarrow w : d$’

3	N	$\rightarrow$	*dog*	:	**dog**
1	V	$\rightarrow$	*dog*	:	$\mathbf{dog}_v$
0	N	$\rightarrow$	*dog*	:	**cat**
2	N	$\rightarrow$	*cat*	:	**cat**
0	N	$\rightarrow$	*cat*	:	**dog**
0	V	$\rightarrow$	*jump*	:	**dog**
1	V	$\rightarrow$	*jump*	:	**jump**

N | *dog* : **dog**   N | *dog* : **dog**

V | *dog* : $\mathbf{dog}_v$ V | *jump* : **jump**

N | *cat* : **cat**   N | *cat* : **cat**

N | *dog* : **dog**

Crude grammars refined via learning

```
for w ∈ Words
    for X ∈ Categories
        for d ∈ Domain
            yield 'X → w : d'
```

3	N	→	*dog*	:	**dog**
1	V	→	*dog*	:	**dog**$_v$
0	N	→	*dog*	:	**cat**
2	N	→	*cat*	:	**cat**
0	N	→	*cat*	:	**dog**
0	V	→	*jump*	:	**dog**
1	V	→	*jump*	:	**jump**

N — *dog* : **dog**   N — *dog* : **dog**

V — *dog* : **dog**$_v$ V — *jump* : **jump**

N — *cat* : **cat**   N — *cat* : **cat**

N — *dog* : **dog**

Crude grammars refined via learning

for $w \in \mathit{Words}$
  for $X \in \mathit{Categories}$
    for $d \in \mathit{Domain}$
      yield '$X \rightarrow w : d$'

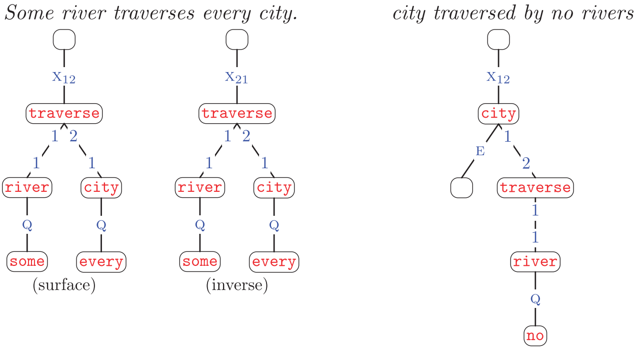
```
/* Sentences */
sentence(S) --> declarative(S), terminator(.) .
sentence(S) --> wh_question(S), terminator(?) .
sentence(S) --> yn_question(S), terminator(?) .
sentence(S) --> imperative(S), terminator(!) .

/* Noun Phrase */
np(np(Agmt,Pronoun,[]),Agmt,NPCase,def,_,Set,Nil) -->
  {is_pp(Set)},
  pers_pron(Pronoun,Agmt,Case),
  {empty(Nil), role(Case,decl,NPCase)}.

/* Prepositional Phrase */
pp(pp(Prep,Arg),Case,Set,Mask) -->
  prep(Prep),
  {prep_case(NPCase)},
  np(Arg,_,NPCase,_,Case,Set,Mask).
```

Crude grammars refined via learning

Liang et al. (2013):



(c) Quantifier scope ambiguity (Q, Q)

(d) Quantification (Q, E)

Additional feature functions

I previously showed some features that correspond to local trees. Those look nicely compositional. However, a smart NLPer will also have features like:

1. How many distinct lexical items are in the sentence?
2. Am I in the c-command domain of a negation?
3. Does this structure contain a specific set of tree fragments?
4. What is the average sentiment of words in this sentence?

The linguist's ideal again

Every child ate a lollipop

$$\forall z\ ((\textbf{child}\ z) \rightarrow (\exists x\ (\textbf{lollipop}\ x) \wedge (\textbf{ate}\ x\ z)))$$

$$\lambda g\ \forall z\ ((\textbf{child}\ z) \rightarrow (g\ z)) \quad \lambda Q\ (Q\ (\lambda y\ (\exists x\ (\textbf{lollipop}\ x) \wedge (\textbf{ate}\ x\ y))$$

$$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z)) \quad \textbf{child}$$

$$\textbf{ate}_S \quad \lambda g\ (\exists x\ (\textbf{lollipop}\ x) \wedge (g\ x))$$

$$\lambda f\ \lambda g\ (\exists x\ (f\ x) \wedge (g\ x)) \quad \textbf{lollipop}$$

Score: $+5$

The linguist's ideal again

Every child ate a lollipop

$$\exists x\ (\mathbf{lollipop}\ x) \wedge (\forall z\ ((\mathbf{child}\ z) \rightarrow (\mathbf{ate}\ x\ z)))$$

$\lambda g\ \forall z\ ((\mathbf{child}\ z) \rightarrow (g\ z))$ $\quad$ $\lambda Q\ (\exists x\ (\mathbf{lollipop}\ x) \wedge (Q\ (\lambda y\ (\mathbf{ate}\ x\ y))))$

$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z))$ $\quad$ $\mathbf{child}$ $\quad$ $\mathbf{ate}_l$ $\quad$ $\lambda g\ (\exists x\ (\mathbf{lollipop}\ x) \wedge (g\ x))$

$\lambda f\ \lambda g\ (\exists x\ (f\ x) \wedge (g\ x))$ $\quad$ $\mathbf{lollipop}$

Score: −2

Recursive deep learning models

Composition with functions or with vectors

Functions

$$[\![A]\!] = \mathbf{fa}([\![B]\!], [\![C]\!])$$

$$[\![B]\!] \qquad [\![C]\!] = \mathbf{fa}([\![D]\!], [\![E]\!])$$

$$[\![D]\!] \qquad [\![E]\!]$$

Vectors

$$f([\![X]\!]^{\mathrm{DL}}, [\![Y]\!]^{\mathrm{DL}}) = \mathbf{g}\left(\left[[\![X]\!]^{\mathrm{DL}}; [\![Y]\!]^{\mathrm{DL}}\right] W\right)$$

$$[\![A]\!]^{\mathrm{DL}} = \mathbf{g}\left(\left[[\![B]\!]^{\mathrm{DL}}; [\![C]\!]^{\mathrm{DL}}\right] W\right)$$

$$[\![B]\!]^{\mathrm{DL}} \qquad [\![C]\!]^{\mathrm{DL}} = \mathbf{g}\left(\left[[\![D]\!]^{\mathrm{DL}}; [\![E]\!]^{\mathrm{DL}}\right] W\right)$$

$$[\![D]\!]^{\mathrm{DL}} \qquad [\![E]\!]^{\mathrm{DL}}$$

Lexicon				
B	−0.42	0.10	0.12	...
D	−0.16	−0.21	0.29	...
E	−0.26	0.31	0.37	...

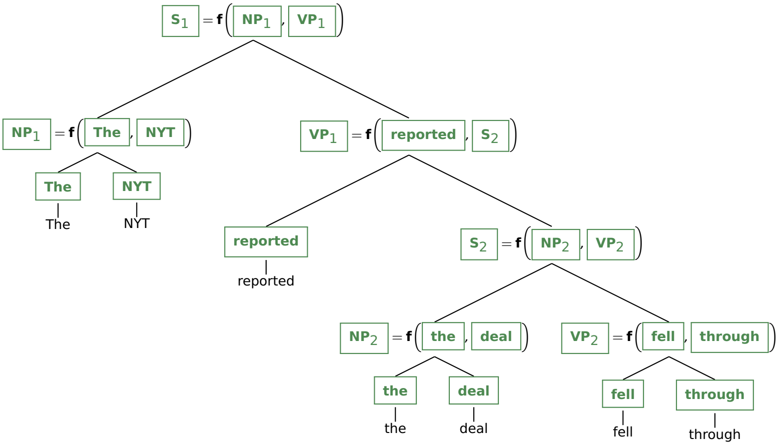
Simple example

$$\tanh\left(\begin{bmatrix}-0.10 & 0.10 & -1.00 & 1.00\end{bmatrix}\begin{bmatrix}0.06 & 0.32\\ -0.14 & -0.53\\ 1.24 & 0.00\\ 0.02 & 1.06\end{bmatrix}\right) = \begin{bmatrix}-0.85 & 0.75\end{bmatrix}$$

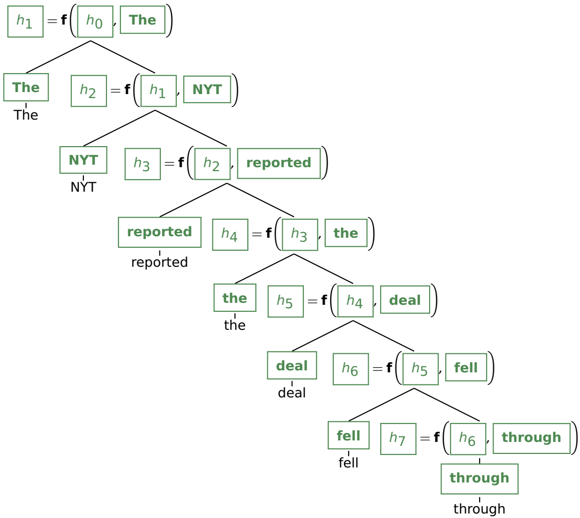
$$\begin{bmatrix}-0.10 & 0.10\end{bmatrix} \qquad \begin{bmatrix}-1.00 & 1.00\end{bmatrix}$$

not terrible

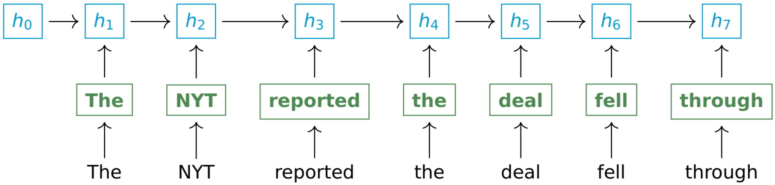
Recursive deep learning models



Recursive deep learning models

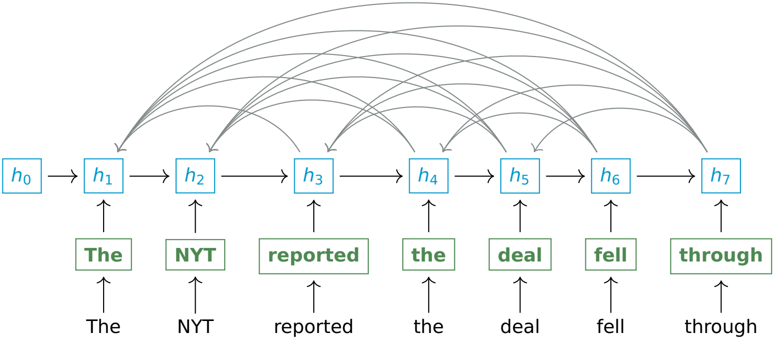


Recursive deep learning models



Recursive deep learning models

All our parses are wrong, but perhaps we can discover the right one(s).



A new perspective on compositionality

A new perspective on compositionality

Partee (1984):

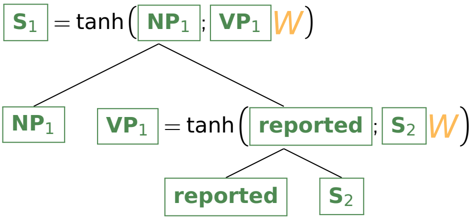
Context-dependence, Ambiguity, and Challenges to Local, Deterministic Compositionality

A new perspective on compositionality

Partee (1984):

Context-dependence, Ambiguity, and Challenges to Local, Deterministic Compositionality

Global parameters creating local lexical effects



A new perspective on compositionality

Partee (1984):

Context-dependence, Ambiguity, and Challenges to Local, Deterministic Compositionality

Global parameters creating local lexical effects

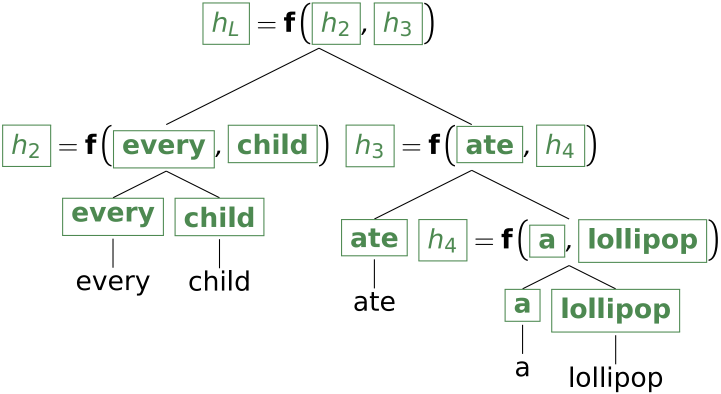
$$\tanh\left(\begin{bmatrix}-0.10 & 0.10 & -1.00 & 1.00\end{bmatrix}\begin{bmatrix}0.06 & 0.32\\ -0.14 & -0.53\\ 1.24 & 0.00\\ 0.02 & 1.06\end{bmatrix}\right) = \begin{bmatrix}-0.85 & 0.75\end{bmatrix}$$

$\begin{bmatrix}-0.10 & 0.10\end{bmatrix}$ $\begin{bmatrix}-1.00 & 1.00\end{bmatrix}$

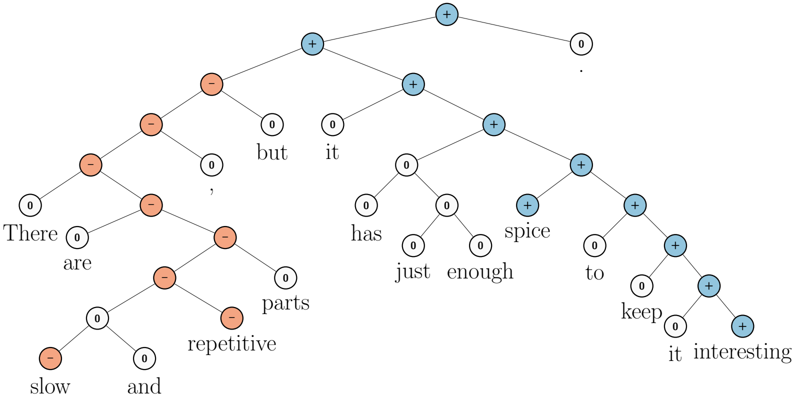
not terrible

The linguist's ideal again

Every child ate a lollipop



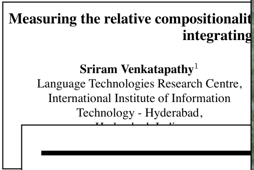
Examples from Socher et al. 2013



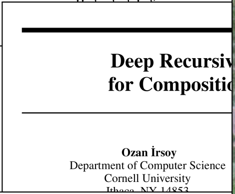
Conclusions

Semantics	Semantic parsing	Deep learning
Recursive	Recursive	Recursive
Symbolic	Symbolic	Not symbolic
Compositional	Partly compositional	Compositional?
Precise	Precise	Not precise
Purely representational	Preferences learned	Preferences learned
Open-class lexicon often neglected	Learned symbolic lexicon	Learned lexical embedding
Rich treatment of functional lexicon	Possible rich treatment of functional lexicon	No functional/open-class distinctions
Sharp sem/prag division	Blurry sem/prag division	No sem/prag division
Not at all scalable	Semi-scalable	Highly scalable
Deep analytic insights	Indirect analytic insights	Often opaque

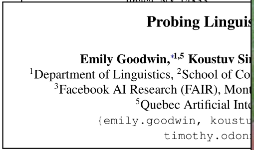
Conclusions


Measuring the relative compositionalit
integrating

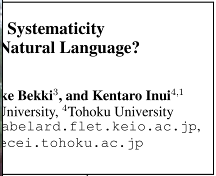
Sriram Venkatapathy[1]
Language Technologies Research Centre,
International Institute of Information
Technology - Hyderabad,


Deep Recursiv
for Compositio

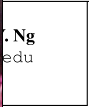
Ozan İrsoy
Department of Computer Science
Cornell University

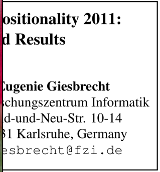

Probing Linguis

Emily Goodwin,[*1,5] Koustuv Si
[1]Department of Linguistics, [2]School of Co
[3]Facebook AI Research (FAIR), Mont
[5]Quebec Artificial Inte
{emily.goodwin, koustu
timothy.odon


Systematicity
Natural Language?

ke Bekki[3], and Kentaro Inui[4,1]
University, [4]Tohoku University
abelard.flet.keio.ac.jp,
ecei.tohoku.ac.jp


. Ng
edu


ositionality 2011:
d Results

ugenie Giesbrecht
chungszentrum Informatik
d-und-Neu-Str. 10-14
1 Karlsruhe, Germany
esbrecht@fzi.de

References I

Biemann, Chris & Eugenie Giesbrecht. 2011. Distributional semantics and compositionality 2011: Shared task description and results. In *Proceedings of the workshop on distributional semantics and compositionality*, 21–28. Portland, Oregon, USA: Association for Computational Linguistics. https://www.aclweb.org/anthology/W11-1304.

Dever, Josh. 1999. Compositionality as methodology. *Linguistics and Philosophy* 22(3). 311–326. doi:10.1023/A:1005410301126.

Dowty, David. 2007. Compositionality as an empirical problem. In Chris Barker & Pauline Jacobson (eds.), *Direct compositionality*, 23–101. Oxford: Oxford University Press.

Goodwin, Emily, Koustuv Sinha & Timothy J. O'Donnell. 2020. Probing linguistic systematicity. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 1958–1969. Online: Association for Computational Linguistics. doi:10.18653/v1/2020.acl-main.177. https://www.aclweb.org/anthology/2020.acl-main.177.

Irsoy, Ozan & Claire Cardie. 2014. Deep recursive neural networks for compositionality in language. *Advances in neural information processing systems* 27. 2096–2104.

Janssen, Theo M. V. 1997. Compositionality. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of logic and language*, 417–473. Cambridge, MA and Amsterdam: MIT Press and North-Holland.

Kazmi, Ali & Francis Jeffry Pelletier. 1998. Is compositionality formally vacuous? *Linguistics and Philosophy* 21(6). 629–633.

Levin, Beth, Lelia Glass & Dan Jurafsky. 2019. Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds. *Linguistics* 57(3). 429–471. doi:10.1515/ling-2019-0013.

Liang, Percy, Michael I. Jordan & Dan Klein. 2013. Learning dependency-based compositional semantics. *Computational Linguistics* 39(2). 389–446. doi:10.1162/COLI_a_00127.

Montague, Richard. 1970. Universal grammar. *Theoria* 36. 373–398. Reprinted in Montague (1974), 222–246.

Montague, Richard. 1974. *Formal philosophy: Selected papers of Richard Montague*. New Haven, CT: Yale University Press.

Partee, Barbara H. 1984. Compositionality. In Fred Landman & Frank Veltman (eds.), *Varieties of formal semantics*, 281–311. Dordrecht: Foris. Reprinted in Barbara H. Partee (2004) *Compositionality in formal semantics*, Oxford: Blackwell 153–181. Page references to the reprinting.

Partee, Barbara H. 1995. Lexical semantics and compositionality. In Lila R. Gleitman & Mark Liberman (eds.), *Invitation to cognitive science*, vol. 1, 311–360. Cambridge, MA: MIT Press.

Partee, Barbara H. 1996. The development of formal semantics in linguistic theory. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 11–38. Oxford: Blackwell.

Socher, Richard, Brody Huval, Christopher D. Manning & Andrew Y. Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1201–1211. Stroudsburg, PA. http://www.aclweb.org/anthology/D12-1110.

References II

Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng & Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on Empirical Methods in Natural Language Processing*, 1631–1642. Stroudsburg, PA: Association for Computational Linguistics. `http://www.aclweb.org/anthology/D13-1170`.

Venkatapathy, Sriram & Aravind Joshi. 2005. Measuring the relative compositionality of verb-noun (V-n) collocations by integrating features. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 899–906. Vancouver, British Columbia, Canada: Association for Computational Linguistics. `https://www.aclweb.org/anthology/H05-1113`.

Warren, David H. D. & Fernando C. N. Pereira. 1982. An efficient easily adaptable system for interpreting natural language queries. *American Journal of Computational Linguistics* 8(3–4). 110–122.

Yanaka, Hitomi, Koji Mineshima, Daisuke Bekki & Kentaro Inui. 2020. Do neural models learn systematicity of monotonicity inference in natural language? In *Proceedings of the 58th annual meeting of the association for computational linguistics*, 6105–6117. Online: Association for Computational Linguistics. doi:10.18653/v1/2020.acl-main.543. `https://www.aclweb.org/anthology/2020.acl-main.543`.

Zadrozny, Wlodek. 1994. From compositional to systematic semantics. *Linguistics and Philosophy* 17(4). 329–342. doi:10.1007/BF00985572.