Compositionality or generalization?

Christopher Potts

Stanford Linguistics

Linguistics Association of Great Britain, Annual Meeting
September 12, 2019

Overview

Overview

My central questions

When linguists seek compositional analyses of linguistic phenomena:

Overview

My central questions

When linguists seek compositional analyses of linguistic phenomena:

- What principles guide their investigations?

Overview

My central questions

When linguists seek compositional analyses of linguistic phenomena:

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

Overview

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

Overview

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,

Overview

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.

Plan

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

Compositional semantics and statistical NLP

Compositional semantics and statistical NLP

Cooper (2012):

There has been a recent intensification of interest in "semantics" in computational linguistics. I write the word in scare quotes because there are very different views of what computational semantics is. Broadly, it divides into the view that word meaning can be modeled in distributional terms and the view that meaning is to be viewed in terms of model theory of the kind employed in formal semantics deriving from the seminal work of Richard Montague (1974).

Compositional semantics and statistical NLP

Cooper (2012):

There has been a recent intensification of interest in “semantics” in computational linguistics. I write the word in scare quotes because there are very different views of what computational semantics is. Broadly, it divides into the view that word meaning can be modeled in distributional terms and the view that meaning is to be viewed in terms of model theory of the kind employed in formal semantics deriving from the seminal work of Richard Montague (1974).

Compositional semantics and statistical NLP

Cooper (2012):

There has been a recent intensification of interest in "semantics" in computational linguistics. I write the word in scare quotes because there are very different views of what computational semantics is. Broadly, it divides into the view that word meaning can be modeled in distributional terms and the view that meaning is to be viewed in terms of model theory of the kind employed in formal semantics deriving from the seminal work of Richard Montague (1974).

Not anymore!

Pater's target article and a historical note

Pater (2019):

When viewed from a sufficient distance, neural network and generative linguistic approaches to cognition overlap considerably: they both aim to provide formally explicit accounts of the mental structures underlying cognitive processes, and they both aim to explain how those structures are learned.

The compositionality principle

Informal statement

Informal statement

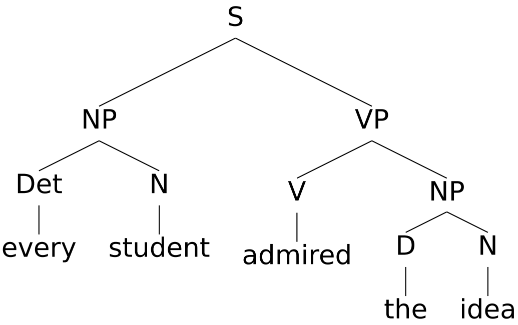
Compositionality

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

Informal statement

Compositionality

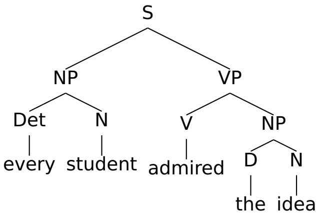
The meaning of a phrase is a function of the meanings of its immediate syntactic constituents and the way they are combined.



The usual motivation

The usual motivation

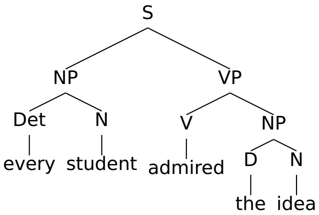
1. Modeling all meaningful units



The usual motivation

1. Modeling all meaningful units

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

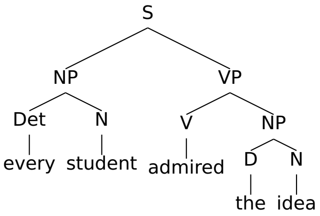


The usual motivation

1. Modeling all meaningful units

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

2. “Infinite” capacity



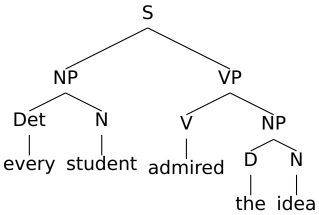
The usual motivation

1. Modeling all meaningful units

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

2. “Infinite” capacity

3. Creativity



The usual motivation

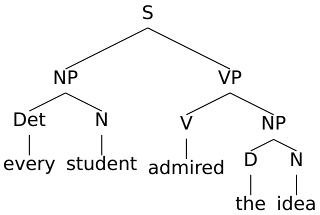
1. Modeling all meaningful units

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

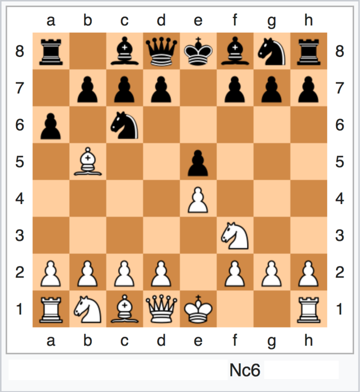
2. “Infinite” capacity

3. Creativity

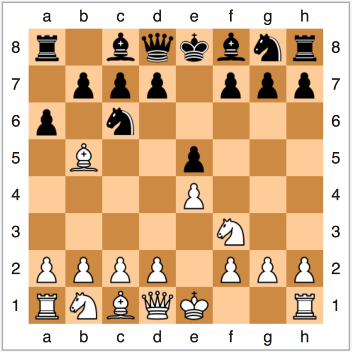
4. Systematicity



Compositionality or systematicity?

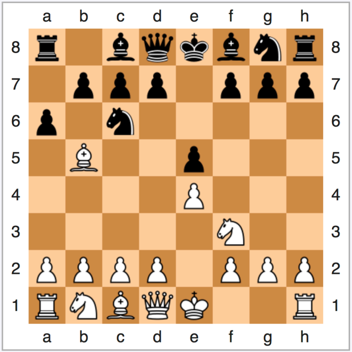


Compositionality or systematicity?



Position after 1.e4 e5 2.Nf3 Nc6 3.Bb5 a6

Compositionality or systematicity?



Position after 1.e4 e5 2.Nf3 Nc6 3.Bb5 a6

Szabó (2012):

The second moral is that – given certain assumptions about meaning in chess notation – we can have productive and systematic understanding of representations even if the system itself is not compositional.

What Frege really said

What Frege really said

Compositionality is often attributed to Frege. But, as Janssen (1997:420) notes, Frege actually only explicitly endorses

What Frege really said

Compositionality is often attributed to Frege. But, as Janssen (1997:420) notes, Frege actually only explicitly endorses

Principle of contextuality

One should ask for the meaning of a word only in the context of a sentence, and not in isolation.

What Frege really said

Compositionality is often attributed to Frege. But, as Janssen (1997:420) notes, Frege actually only explicitly endorses

Principle of contextuality

One should ask for the meaning of a word only in the context of a sentence, and not in isolation.

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

Montague: Unconstrained compositionality

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.

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 [. . .].

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.

Propositional logic semantics as homomorphism

$$\llbracket p \rrbracket \in \{\emptyset, \{\emptyset\}\} \text{ for all propositional letters } p$$
$$\llbracket \neg\varphi \rrbracket = \{\emptyset, \{\emptyset\}\} - \llbracket \varphi \rrbracket$$
$$\llbracket \varphi \vee \psi \rrbracket = \llbracket \varphi \rrbracket \cup \llbracket \psi \rrbracket$$
$$\llbracket \varphi \wedge \psi \rrbracket = \llbracket \varphi \rrbracket \cap \llbracket \psi \rrbracket$$

Formal results

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.

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:

- $[\![A]\!] = [\![B]\!]$
- $[\![C.A]\!] \neq [\![C.B]\!]$

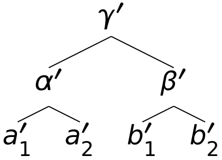
Zadrozny showed how to create a kind of syntactic layer where compositionality holds. A similar argument is made by Dever 1999.

Dowty's context-free compositionality

Dowty’s context-free compositionality

Dowty (2007):

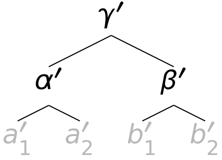
When a rule f combines $\alpha, \beta(\ldots)$ to form γ, the corresponding semantic rule g that produces the meaning γ' of γ, from α' and β', may depend only on α' “as a whole”, it may not depend on any meanings from which α' was formed compositionally by earlier derivational steps (similarly for β).



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

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

Compositionality as methodology

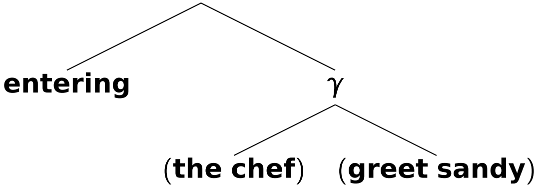
Compositionality as methodology

Janssen (1997:461)

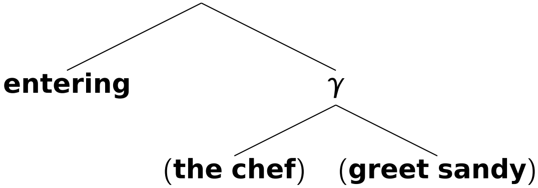
Compositionality is not a formal restriction on what can be achieved, but a methodology on how to proceed.

Example: Subjectless predicational adjuncts

Example: Subjectless predicational adjuncts

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

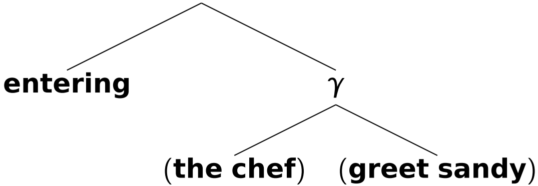
Example: Subjectless predicational adjuncts

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

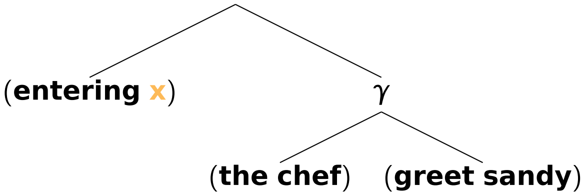
Pragmatic constraint

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

Example: Subjectless predicational adjuncts

Pragmatic constraint

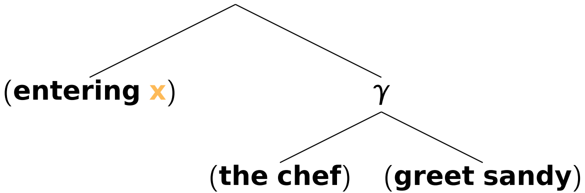
The free variable in a subjectless predicational adjuncts should refer to a discourse topic.

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: Exceptives

Example: Exceptives

Assumptions

Example: Exceptives

Assumptions

1. Exceptives adjoin only to universal quantifiers:
   a. Every Muppet except Kermit danced.
   b. No Muppet except Kermit danced.
   c. *Most Muppets except Kermit danced.

Example: Exceptives

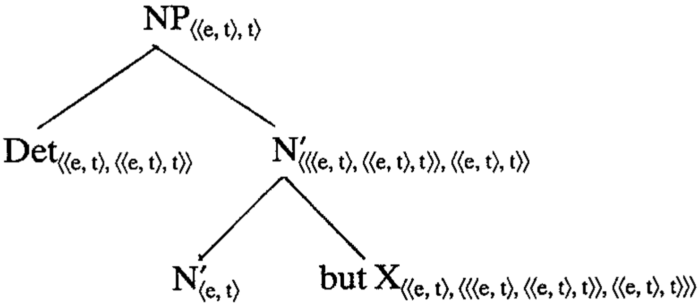
Assumptions

1. Exceptives adjoin only to universal quantifiers:
   a. Every Muppet except Kermit danced.
   b. No Muppet except Kermit danced.
   c. *Most Muppets except Kermit danced.

2. The exceptive's determiner affects VP entailments:
   a. Every Muppet except Kermit danced.
      ⇒ Kermit didn't dance.
   b. No Muppet except Kermit danced.
      ⇒ Kermit did dance.

Example: Exceptives

von Fintel (1993):



Example: Exceptives

Assumptions challenged

Example: Exceptives

Assumptions challenged

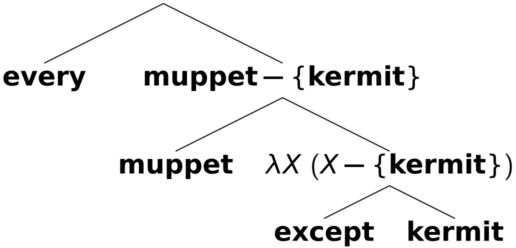
1. Exceptives adjoin to lots of phrases!
   a. Does poetry matter? Few but other poets may read it. (Horn 2005)
   b. She didn’t eat almost anything except pills. (Horn 2005)
   c. “an embarrassment for most Americans except his base.”

Example: Exceptives

Assumptions challenged

1. Exceptives adjoin to lots of phrases!
   a. Does poetry matter? Few but other poets may read it. (Horn 2005)
   b. She didn't eat almost anything except pills. (Horn 2005)
   c. "an embarrassment for most Americans except his base."
2. The VP entailments are merely implicated!
   a. Well, we can't find Karl, but we've verified that everyone except Karl has an alibi, so let's find out whether he does too. (Hoeksema 1996)
   b. "All of you are moving on [to Spanish 102] except for Jeff. Turns out you – pause for dramatic effect – will be seeing me next semester. In Spanish 102. Because he passed, you know, and I'm the only Spanish teacher." (*Community* 1:12).

Example: Exceptives



Example: Exceptives

- **every**
- **muppet** $-$ {**kermit**}
  - **muppet**
  - $\lambda X\ (X - \{$**kermit**$\})$
    - **except**
    - **kermit**

See also Higginbotham 1986 and
Nadathur & Lassiter (2014) on *unless*

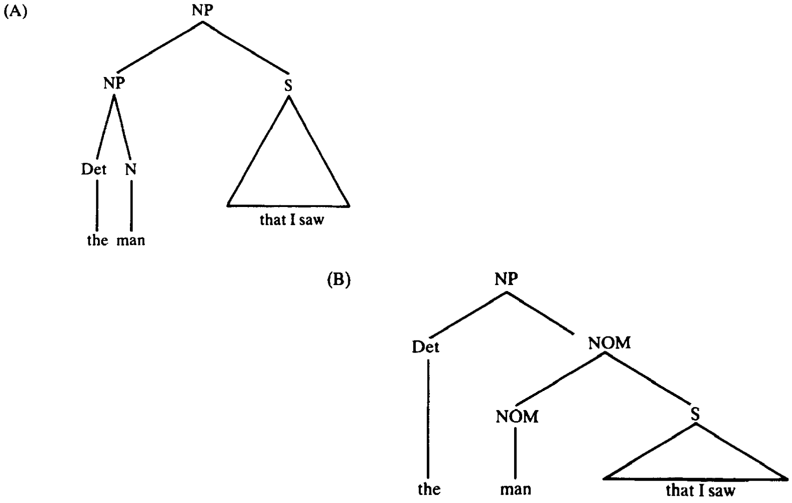
Example: Restrictive relative clauses

Example: Restrictive relative clauses

Bach & Cooper (1978):

Example: Restrictive relative clauses

Bach & Cooper (1978):



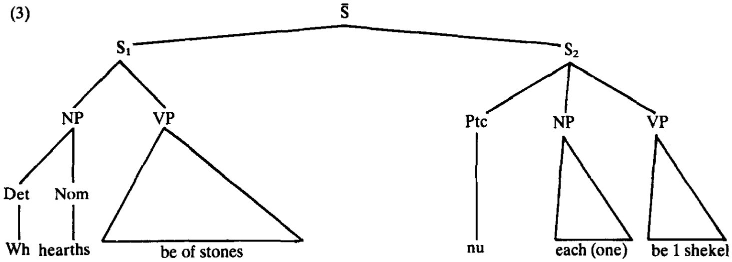
Example: Restrictive relative clauses

Bach & Cooper (1978):

(1) And every hearth which is made of stones costs 1 shekel.

(2) *ŠA* NA$_4$. HI. A.-ia kuieš GUNNI . MEŠ nu kuišša 1 GÌN
gen. stone-pl. -and which hearth-pl. ptc. each (one) 1 shekel

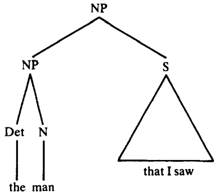
(3)



Example: Restrictive relative clauses

Bach & Cooper (1978):

(A)

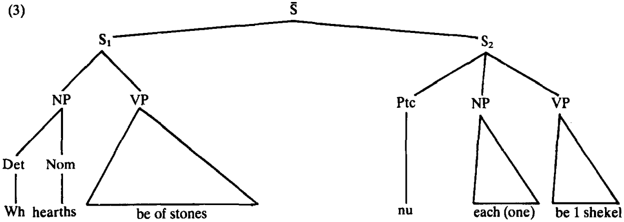


$$\textbf{the}\ (\lambda x\ ((\textbf{man}\ x) \wedge (\textbf{f}\ x)))$$

$$\textbf{the} \qquad \lambda x\ ((\textbf{man}\ x) \wedge (\textbf{f}\ x))$$

$$\textbf{man} \qquad \textbf{f}$$

(3)



Things "in the wrong place" (Partee 1984)

Things “in the wrong place” (Partee 1984)

1. An occasional sailor walked by. (Bolinger 1967)

Things “in the wrong place” (Partee 1984)

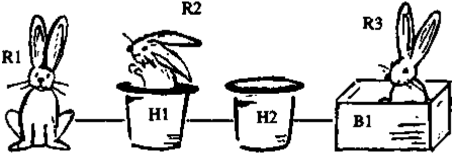
1. An occasional sailor walked by. (Bolinger 1967)
2. Sandy enjoyed a quiet cup of tea. (Hall 1973)

Things "in the wrong place" (Partee 1984)

1. An occasional sailor walked by. (Bolinger 1967)
2. Sandy enjoyed a quiet cup of tea. (Hall 1973)
3. The same waiter served everyone. (Barker 2007)

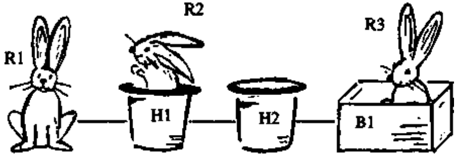
Things “in the wrong place” (Partee 1984)

1. An occasional sailor walked by. (Bolinger 1967)
2. Sandy enjoyed a quiet cup of tea. (Hall 1973)
3. The same waiter served everyone. (Barker 2007)
4. the rabbit in the hat (Haddock 1987)



Things “in the wrong place” (Partee 1984)

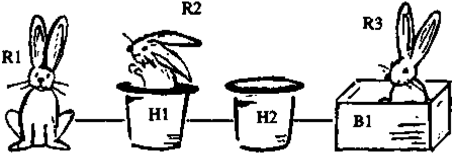
1. An occasional sailor walked by. (Bolinger 1967)
2. Sandy enjoyed a quiet cup of tea. (Hall 1973)
3. The same waiter served everyone. (Barker 2007)
4. the rabbit in the hat (Haddock 1987)



“Any compositional accounts of NP semantics [. . .] would judge (1) to be infelicitous.”

Things “in the wrong place” (Partee 1984)

1. An occasional sailor walked by. (Bolinger 1967)
2. Sandy enjoyed a quiet cup of tea. (Hall 1973)
3. The same waiter served everyone. (Barker 2007)
4. the rabbit in the hat (Haddock 1987)



“Any compositional accounts of NP semantics [. . .] would judge (1) to be infelicitous.”

Bumford (2017): these cases show us that definites introduce new discourse referents and constrain them, and these operations can take scope separately.

Example: Compounds

Example: Compounds

Partee (1995:341):

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

Example: Compounds

Partee (1995:341):

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

Levin et al.'s (2018) novel compounds experiment:

Modifier	**Head**	**Example**	**Top inference**	**Rate**
Artifact	Artifact	stew skillet	Event	97.3%
Natural kind	Artifact	stream wheel	Event	96.4%
Artifact	Natural kind	stew chickpea	Event	76.9%
Natural kind	Natural kind	stream vegetable	Essence	92.3%

Example: Compounds

Partee (1995:341):

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

Levin et al.'s (2018) novel compounds experiment:

Modifier	Head	Example	Top inference	Rate
Artifact	Artifact	stew skillet	Event	97.3%
Natural kind	Artifact	stream wheel	Event	96.4%
Artifact	Natural kind	stew chickpea	Event	76.9%
Natural kind	Natural kind	stream vegetable	Essence	92.3%

(Why is there consensus that compounds aren't compositional but no real concern about *flat tire/beer/note* or *throw a ball/party/fight* (though see Keenan 1974; Hirst 1987; Partee 1984)?)

Other trends

Other trends

1. Enthusiastically non-compositional theories (e.g., DRT; Kamp 1981) are expected to have compositional variants (e.g., DPL; Groenendijk & Stokhof 1991).

Other trends

1. Enthusiastically non-compositional theories (e.g., DRT; Kamp 1981) are expected to have compositional variants (e.g., DPL; Groenendijk & Stokhof 1991).

2. Contextual ambiguity resolution is no longer taken to challenge compositionality:
   a. I saw a crane.
   b. The crane flew away.
   c. The crane lifted the beam.

Other trends

1. Enthusiastically non-compositional theories (e.g., DRT; Kamp 1981) are expected to have compositional variants (e.g., DPL; Groenendijk & Stokhof 1991).

2. Contextual ambiguity resolution is no longer taken to challenge compositionality:
   a. I saw a crane.
   b. The crane flew away.
   c. The crane lifted the beam.

3. Pragmatic free variables are no longer taken to be a challenge to compositionality:
   a. Being a master of disguise, Bill would fool anyone.
   b. Wearing his new outfit, Bill would fool anyone.

Semantic parsing

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

The semanticist's ideal

Every student attended a lecture

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

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

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

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

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

The semanticist's ideal

Every student attended a lecture

$$\exists x\ (\mathbf{lecture}\ x) \land (\forall z\ ((\mathbf{student}\ z) \rightarrow (\mathbf{attended}\ x\ z)))$$

$\lambda g\ \forall z\ ((\mathbf{student}\ z) \rightarrow (g\ z))$ $\quad$ $\lambda Q\ (\exists x\ (\mathbf{lecture}\ x) \land (Q\ (\lambda y\ (\mathbf{attended}\ x\ y))))$

$\lambda f\ \lambda g\ \forall z\ ((f\ z) \rightarrow (g\ z))$ $\quad$ $\mathbf{student}$ $\quad$ $\mathbf{attended}_l$ $\quad$ $\lambda g\ (\exists x\ (\mathbf{lecture}\ x) \land (g\ x))$

$\lambda f\ \lambda g\ (\exists x\ (f\ x) \land (g\ x))$ $\quad$ $\mathbf{lecture}$

The semanticist's ideal

Every child ate a lollipop

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

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

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

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

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

The semanticist's ideal

Every child ate a lollipop

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

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

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

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

The semanticist's ideal

Every child ate a lollipop

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

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

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

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

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

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 $w \in \mathit{Words}$
    for $X \in \mathit{Categories}$
        for $d \in \mathit{Domain}$
            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

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

for $w \in \mathit{Words}$
    for $X \in \mathit{Categories}$
        for $d \in \mathit{Domain}$
            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

$$\begin{array}{l} \textbf{for } w \in \textit{Words} \\ \quad \textbf{for } X \in \textit{Categories} \\ \quad\quad \textbf{for } d \in \textit{Domain} \\ \quad\quad\quad \textbf{yield } `X \rightarrow w : d\text{'} \end{array}$$

N
|
dog : **dog**

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

1	N	$\rightarrow$	*dog*	:	**dog**
1	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

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

2	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**

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

N
|
cat : **cat**

N
|
dog : **dog**

Crude grammars refined via learning

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

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**

N — *dog* : **dog**

N — *dog* : **dog**

V — *dog* : $\mathbf{dog}_v$

V — *jump* : **jump**

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 ‘$X \rightarrow w : d$’

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

3	N	→	*dog*	:	**dog**
1	V	→	*dog*	:	$\mathbf{dog}_V$
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 : $\mathbf{dog}_V$

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'
```

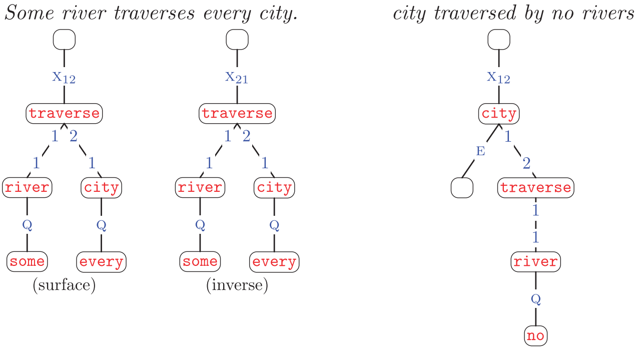
```
/* Sentences */
sentence(S) --> declarative(S), terminator(.) .
sentence(S) --> wh_question(S), terminator(?) .
sentence(S) --> yn_question(S), terminator(?) .
sentence(S) --> imperative(S), terminator(!) .

/* 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:

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?

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?

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?

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?

Comparison with traditional semantic theory

Comparison with traditional semantic theory

1. Both are symbolic.

Comparison with traditional semantic theory

1. Both are symbolic.
2. Both involve compositional grammars.

Comparison with traditional semantic theory

1. Both are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
    - Semantic parsing grammars process examples holistically, greedily extracting useful information. Truly Fregean compositionality?
    - Linguistic grammars are narrowly compositional.

Comparison with traditional semantic theory

1. Both are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
   - Semantic parsing grammars process examples holistically, greedily extracting useful information. Truly Fregean compositionality?
   - Linguistic grammars are narrowly compositional.
4. Overgeneration:
   - Semantic parsing grammars generate lots of junk.
   - Linguistic grammars ought not to.

Comparison with traditional semantic theory

1. Both are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
   - Semantic parsing grammars process examples holistically, greedily extracting useful information. Truly Fregean compositionality?
   - Linguistic grammars are narrowly compositional.
4. Overgeneration:
   - Semantic parsing grammars generate lots of junk.
   - Linguistic grammars ought not to.
5. Ambiguity resolution:
   - Semantic parsing grammars make predictions about preferences for ambiguity resolution.
   - Linguistic grammars limit themselves to representation.

The linguist's ideal again

Every child ate a lollipop

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

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

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

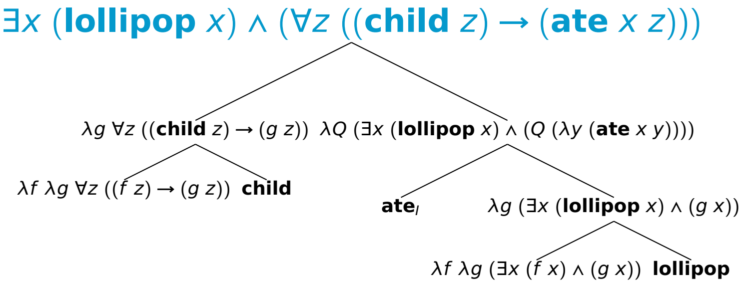
$$\mathbf{ate}_S \qquad \lambda g\ (\exists x\ (\mathbf{lollipop}\ x) \wedge (g\ x))$$

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

Score: $+5$

The linguist's ideal again

Every child ate a lollipop



Score: −2

Recursive deep learning models

Composition with functions or with vectors

Composition with functions or with vectors

Functions

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

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

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

Composition with functions or with vectors

Functions

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

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

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

Vectors

Composition with functions or with vectors

Functions

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

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

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

Vectors

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

Composition with functions or with vectors

Functions

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

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

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

Vectors

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

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

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

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

Composition with functions or with vectors

Functions

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

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

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

Vectors

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

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

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

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

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

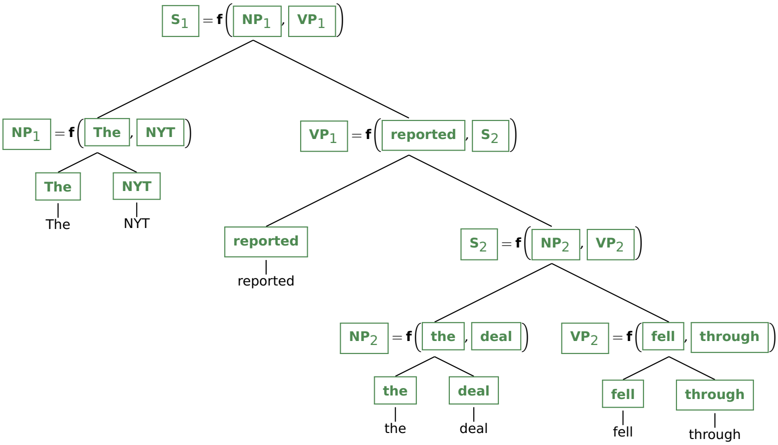
Simple example

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

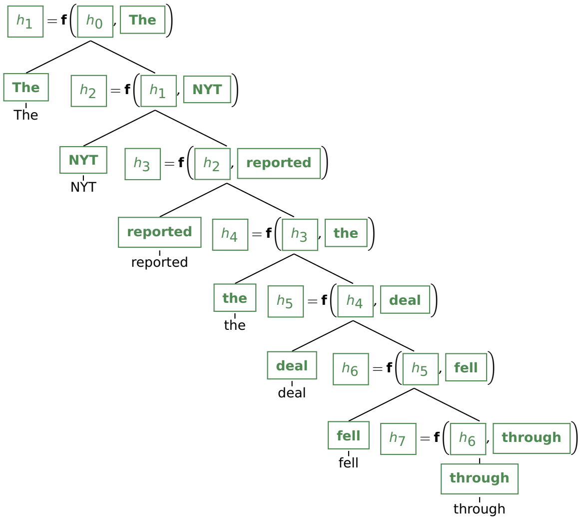
$\begin{bmatrix}-0.10 & 0.10\end{bmatrix}$ $\begin{bmatrix}-1.00 & 1.00\end{bmatrix}$

not terrible

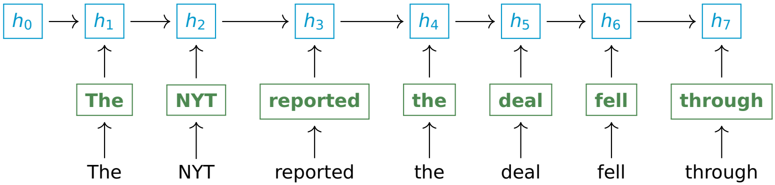
Recursive deep learning models



Recursive deep learning models

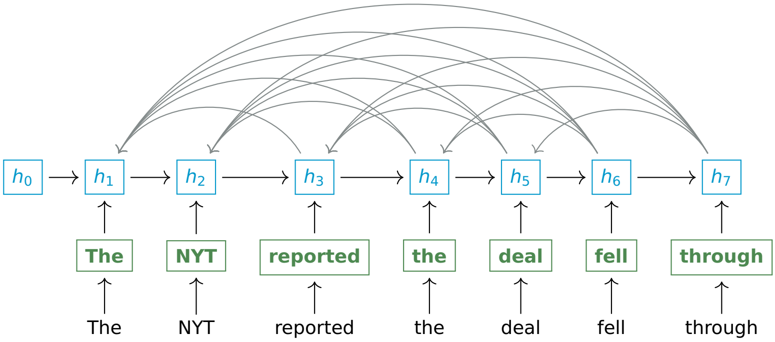


Recursive deep learning models



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

$$\boxed{\mathbf{S}_1} = \tanh\left(\boxed{\mathbf{NP}_1};\boxed{\mathbf{VP}_1}\,W\right)$$

$$\boxed{\mathbf{NP}_1} \qquad \boxed{\mathbf{VP}_1} = \tanh\left(\boxed{\textbf{reported}};\boxed{\mathbf{S}_2}\,W\right)$$

$$\boxed{\textbf{reported}} \qquad \boxed{\mathbf{S}_2}$$

A new perspective on compositionality

Partee (1984):

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

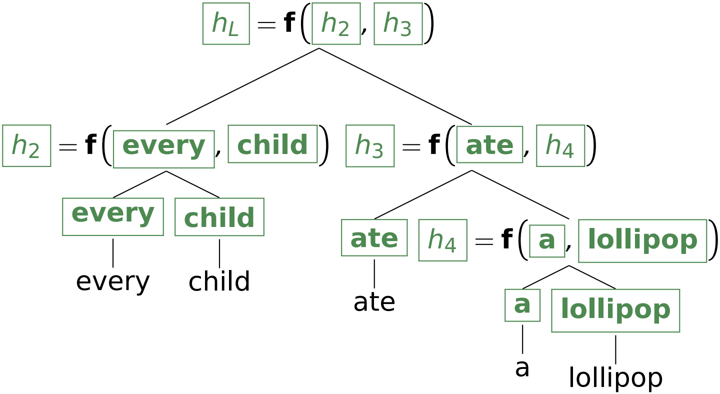
Global parameters creating local lexical effects

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

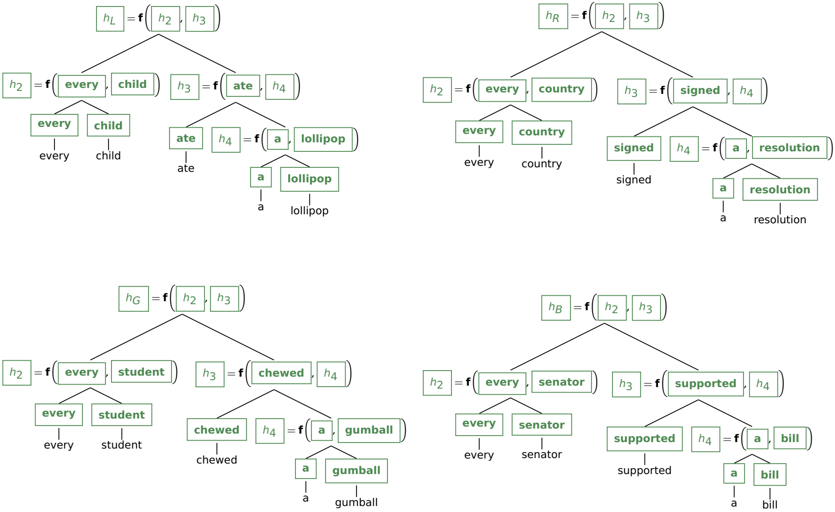
$$\begin{array}{cc} \begin{bmatrix}-0.10 & 0.10\end{bmatrix} & \begin{bmatrix}-1.00 & 1.00\end{bmatrix} \\ | & | \\ \text{not} & \text{terrible} \end{array}$$

The linguist's ideal again

Every child ate a lollipop

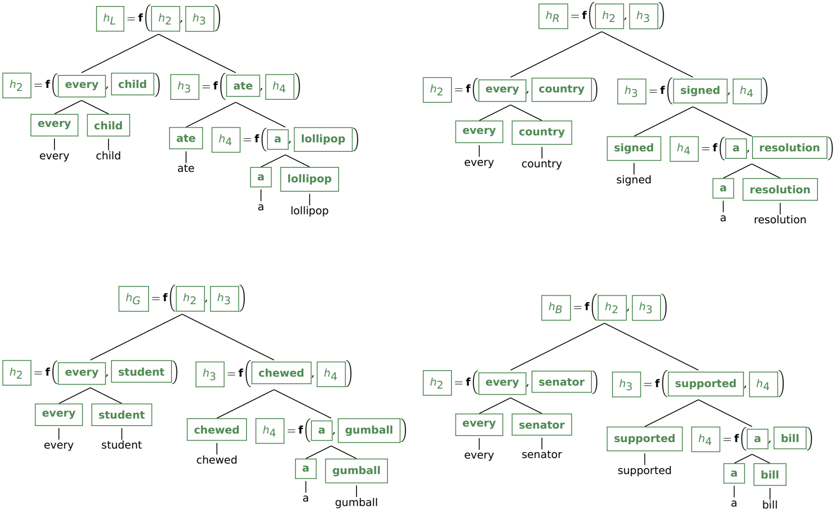


The linguist's ideal again



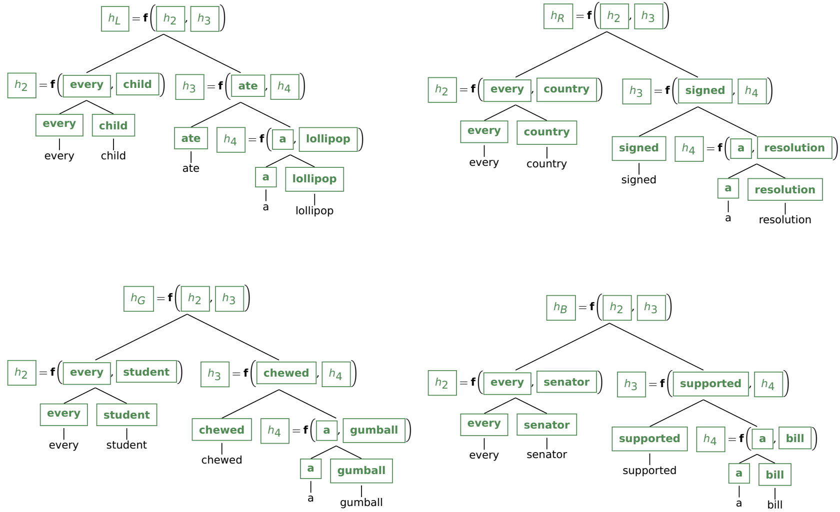
The linguist's ideal again

$$\textbf{sim}\left(\boxed{h_L}, \boxed{h_G}\right) > \textbf{sim}\left(\boxed{h_L}, \boxed{h_R}\right)$$

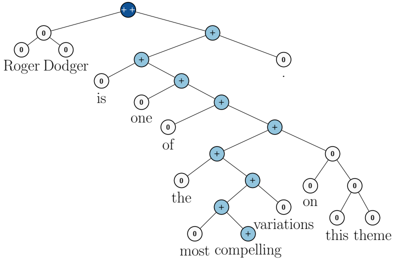


The linguist's ideal again

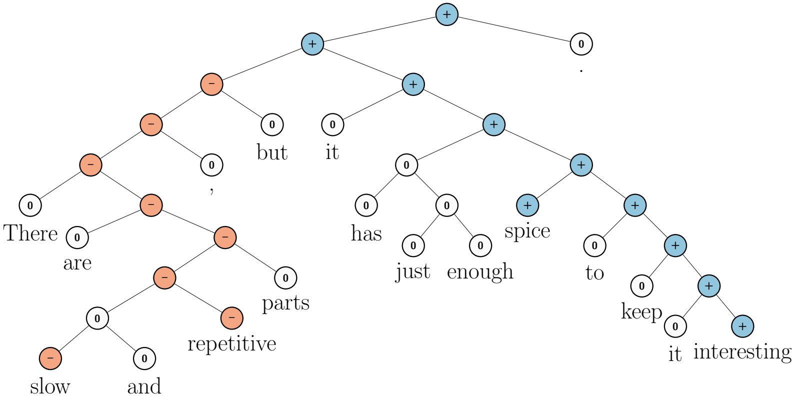
$$\textbf{sim}\left(\boxed{h_L}, \boxed{h_G}\right) > \textbf{sim}\left(\boxed{h_L}, \boxed{h_R}\right) \text{ or } \textbf{g}\left(\boxed{h_L}, \boxed{\textbf{many lollipops eaten}}\right) = \textbf{entail}$$



Examples from Socher et al. 2013



Examples from Socher et al. 2013



Monotonicity and semantic *precision*

Monotonicity and semantic *precision*

Kim smoked.	Kim didn't smoke.
↑	↓
Kim smoked cigars.	Kim didn't smoke cigars.

Monotonicity and semantic *precision*

Kim smoked. $\uparrow$ Kim smoked cigars.

Kim didn't smoke. $\downarrow$ Kim didn't smoke cigars.

A student smoked.

$\nearrow$ **A** Swedish student smoked.

$\nwarrow$ **A** student smoked cigars.

Monotonicity and semantic *precision*

Kim smoked.
↑
Kim smoked cigars.

Kim didn't smoke.
↓
Kim didn't smoke cigars.

A student smoked.
↗ ↖
A Swedish student smoked. **A** student smoked cigars.

No student smoked.
↙ ↘
No Swedish student smoked. **No** student smoked cigars.

Monotonicity and semantic *precision*

Kim smoked.
↑
Kim smoked cigars.

Kim didn’t smoke.
↓
Kim didn’t smoke cigars.

A student smoked.
↗ ↖
A Swedish student smoked. **A** student smoked cigars.

No student smoked.
↙ ↘
No Swedish student smoked. **No** student smoked cigars.

Every student smoked.
↙ ↖
Every Swedish student smoked. **Every** student smoked cigars.

Monotonicity and semantic *precision*

Kim smoked. ↑ Kim smoked cigars.

Kim didn't smoke. ↓ Kim didn't smoke cigars.

A student smoked.

↗ ↖

A Swedish student smoked. **A** student smoked cigars.

No student smoked.

↙ ↘

No Swedish student smoked. **No** student smoked cigars.

Every student smoked.

↙ ↖

Every Swedish student smoked. **Every** student smoked cigars.

Most students smoked.

— ↖

Most Swedish students smoked. **Most** students smoked cigars.

Monotonicity and semantic *precision*

Kim smoked.	Kim didn’t smoke.
↑	↓
Kim smoked cigars.	Kim didn’t smoke cigars.

A student smoked.

↗ ↖

A Swedish student smoked. **A** student smoked cigars.

No student smoked.

↙ ↘

No Swedish student smoked. **No** student smoked cigars.

Every student smoked.

↙ ↖

Every Swedish student smoked. **Every** student smoked cigars.

Most students smoked.

— ↖

Most Swedish students smoked. **Most** students smoked cigars.

(Bowman 2016)

Comparison with traditional semantic theory

Comparison with traditional semantic theory

1. Symbolic:
   - Deep learning grammars are not symbolic (sub-symbolic?).
   - Linguistic grammars are symbolic.

Comparison with traditional semantic theory

1. Symbolic:
   - Deep learning grammars are not symbolic (sub-symbolic?).
   - Linguistic grammars are symbolic.
2. Both involve compositional grammars.

Comparison with traditional semantic theory

1. Symbolic:
   - Deep learning grammars are not symbolic (sub-symbolic?).
   - Linguistic grammars are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
   - Deep learning systems learn global parameters that shape how composition proceeds. Truly Fregean compositionality?
   - Linguistic grammars are narrowly compositional.

Comparison with traditional semantic theory

1. Symbolic:
   - Deep learning grammars are not symbolic (sub-symbolic?).
   - Linguistic grammars are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
   - Deep learning systems learn global parameters that shape how composition proceeds. Truly Fregean compositionality?
   - Linguistic grammars are narrowly compositional.
4. Precision:
   - Current deep learning are not precise.
   - Linguistic grammars are extremely precise.

Comparison with traditional semantic theory

1. Symbolic:
   - Deep learning grammars are not symbolic (sub-symbolic?).
   - Linguistic grammars are symbolic.
2. Both involve compositional grammars.
3. Interpretation:
   - Deep learning systems learn global parameters that shape how composition proceeds. Truly Fregean compositionality?
   - Linguistic grammars are narrowly compositional.
4. Precision:
   - Current deep learning are not precise.
   - Linguistic grammars are extremely precise.
5. Ambiguity resolution:
   - Deep learning grammars make predictions about preferences for ambiguity resolution.
   - Linguistic grammars limit themselves to representation.

Conclusions

Semantics	Semantic parsing	Deep learning

Conclusions

Semantics	Semantic parsing	Deep learning
Recursive	Recursive	Recursive

Conclusions

Semantics	Semantic parsing	Deep learning
Recursive	Recursive	Recursive
Symbolic	Symbolic	Not symbolic

Conclusions

Semantics	Semantic parsing	Deep learning
Recursive	Recursive	Recursive
Symbolic	Symbolic	Not symbolic
Compositional	Partly compositional	Compositional?

Conclusions

Semantics	**Semantic parsing**	**Deep learning**
Recursive	Recursive	Recursive
Symbolic	Symbolic	Not symbolic
Compositional	Partly compositional	Compositional?
Precise	Precise	Not precise

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

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

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

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

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

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

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

Thanks!

References I

Bach, Emmon & Robin Cooper. 1978. The NP–S analysis of relative clauses and compositional semantics. *Linguistics and Philosophy* 2(1). 145–149.

Barker, Chris. 2007. Parasitic scope. *Linguistics and Philosophy* 30(4). 407–444.

Bolinger, Dwight. 1967. Adjectives in english: Attribution and predication. *Lingua* 18(1). 1–34.

Bowman, Samuel R. 2016. *Modeling natural language semantics in learned representations*. Stanford, CA: Stanford University dissertation.

Bumford, Dylan. 2017. Split-scope definites: Relative superlatives and Haddock descriptions. *Linguistics and Philosophy* 40(6). 549–593. doi:10.1007/s10988-017-9210-2.

Cooper, Robin. 2012. Book review: Computational semantics with functional programming Jan van Eijck and Christina Unger. *Computational Linguistics* 38(2). 447–449. doi:10.1162/COLI_r_00103.

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.

von Fintel, Kai. 1993. Exceptive constructions. *Natural Language Semantics* 1(2). 123–148.

Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14(1). 39–100.

Haddock, Nicolas J. 1987. Incremental interpretation and combinatory categorial grammar. In *Proceedings of the 10th international joint conference on artificial intelligence*, 661–663.

Hall, Robert A., Jr. 1973. The transferred epithet in P. G. Wodehouse. *Linguistic Inquiry* 4(1). 92–94.

Higginbotham, James. 1986. Linguistic theory and davidson's program in semantics. In Ernest Lepore (ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson*, 29–48. Oxford: Basil Blackwell.

Hirst, Graeme. 1987. *Semantic interpretation and the resolution of ambiguity*. Cambridge: Cambridge University Press.

Hoeksema, Jacob. 1996. The semantics of exception phrases. In Jaap van der Does & Jan van Eijck (eds.), *Quantifiers, logic, and language*, 145–177. Stanford, CA: CSLI.

Horn, Laurence R. 2005. Diagnosing a diagnostic: Revisiting the quantifier constraint on exceptives. NYU Polarity Worskhop.

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.

Kamp, Hans. 1981. A theory of truth and discourse representation. In Jeroen Groenendijk, Theo M. V. Janssen & Martin Stockhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Centre.

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

Keenan, Edward L. 1974. The functional principle: Generalizing the notion 'subject of'. In Michael W. La Galy, Robert A. Fox & Anthony Bruck (eds.), *Papers CLS 10*, 298–309. Chicago: Chicago Linguistic Society.

Levin, Beth, Lelia Glass & Dan Jurafsky. 2018. Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds. *Linguistics* .

References II

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.

Nadathur, Prerna & Daniel Lassiter. 2014. Unless: An experimental approach. In *Proceedings of sinn und bedeutung 19*, Göttingen.

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.

Pater, Joe. 2019. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95(1). e41–e74.

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`.

Szabó, Zoltán Gendler. 2012. Compositionality. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*, CSLI winter 2012 edn. `http://plato.stanford.edu/archives/win2012/entries/compositionality/`.

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.

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