Recap: Distributional semantics
What I’ll Cover...

- Recap: Distributional semantics
- Compositionality vs. context dependence
What I’ll Cover...

- Recap: Distributional semantics
- Compositionality vs. context dependence
- Composition in vector space models
What I’ll Cover...

- Recap: Distributional semantics
- Compositionality vs. context dependence
- Composition in vector space models
- Truth-conditional distributional semantics (my work)
“Similar words appear in similar contexts”
(For history and discussion, see: Sahlgren, 2008, “The distributional hypothesis”)
“Similar words appear in similar contexts”
(For history and discussion, see: Sahlgren, 2008, “The distributional hypothesis”)

“The contexts in which an expression appears give us information about its meaning”
(Emerson, 2020, “What are the Goals of Distributional Semantics?”)
... being hurt by another horse especially if some rider ...
... beaten by a better horse at the distance on ...
... from these studies that horses reared with other horses reared with other horses in a free and ...
... ‘Is that all your horse gets to eat?’ in ...
... cache of cattle and horse bones, while from the ...
... was a sterling good horse, especially at Ascot, but ...
... way as a domestic horse that it is stabled ...
... 1790 – that is, one horse or two cows for ...
... as coarse as a horse’s tail straying from ...
...
... being hurt by another horse especially if some rider...
... beaten by a better horse at the distance on...
... from these studies that horses reared with other horses...
... horses reared with other horses in a free and...
... ‘Is that all your horse gets to eat?’ in...
... cache of cattle and horse bones, while from the...
... was a sterling good horse, especially at Ascot, but...
... way as a domestic horse, that it is stabled...
... 1790 – that is, one horse or two cows for...
... as coarse as a horse’s tail straying from...
... being hurt by another horse especially if some rider ... 
... beaten by a better horse at the distance on ...
... from these studies that horses reared with other horses ... 
... horses reared with other horses in a free and ...
... ‘Is that all your horse gets to eat?’ in ...
... cache of cattle and horse bones, while from the ...
... was a sterling good horse, especially at Ascot, but ...
... way as a domestic horse that it is stabled ...
... 1790 – that is, one horse or two cows for ...
... as coarse as a horse’s tail straying from ...
... being hurt by another horse especially if some rider ...
... beaten by a better horse at the distance on ...
... from these studies that horses reared with other horses ...
... horses reared with other horses in a free and ...
... ‘Is that all your horse gets to eat?’ in ...
... cache of cattle and horse bones, while from the ...
... was a sterling good horse, especially at Ascot, but ...
... way as a domestic horse, that it is stabled ...
... 1790 – that is, one horse or two cows for ...
... as coarse as a horse’s tail straying from ...
The context of a word gives us information about its meaning
The context of a word gives us information about its meaning.

What should the model learn?

How can the model learn it?
The context of a word gives us information about its meaning

What should the model learn?
- Mainstream NLP: vectors

How can the model learn it?
The context of a word gives us information about its meaning.

What should the model learn?
- Mainstream NLP: vectors

How can the model learn it?
- Skip-gram, Transformers, ...
Principle of Compositionality: “The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined” (Partee, 1984)
Compositionality

- Principle of Compositionality: “The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined” (Partee, 1984)

- (What is allowed in a meaning representation?)

- (What is allowed in a composition function?)
Productivity: new combinations are immediately understandable
Compositionality

- Productivity: new combinations are immediately understandable
- Productivity requires compositionality
Context Dependence

- The meaning of one part depends on another?
  - \{big, small\} \{elephant, dog, ant, ...\}
The meaning of one part depends on another?

- {big, small} {elephant, dog, ant, ...}
- Productive! Compositional!
Context Dependence

- The meaning of one part depends on another?
  - \{big, small\} \{elephant, dog, ant, ...\}
  - Productive! Compositional!
  - Lexical semantics harder than it might seem...
Context Dependence

- The meaning of one part depends on another?
  - \{big, small\} \{elephant, dog, ant, \ldots\}
  - Productive! Compositional!
  - Lexical semantics harder than it might seem...

- Idioms not compositional (or only semi-compositional):
  - Big Bang (early universe)
  - Big Apple (New York)
Contextualisation vs. Composition

- Contextualised representations
  - One per token (e.g. BERT vectors)
Contextualisation vs. Composition

- Contextualised representations
  - One per token (e.g. BERT vectors)

- Compositional representations
  - One for whole expression (e.g. semantic graph)
Vector Composition

- Goal: derive one vector for an expression
Vector Composition

- Goal: derive one vector for an expression
- Option 1: explicit operation
- Option 2: neural network
Vector Operations

- Vector addition surprisingly effective
Vector Operations

- Vector addition surprisingly effective
  - But insensitive to word order

- (Multiplication is addition of log vectors)

- Tensors have compositional structure
  - (Coecke et al., 2010; Baroni et al., 2013)

- ... but high-order tensors are hard to learn
Vector Operations

- Vector addition surprisingly effective
  - But insensitive to word order
- (Multiplication is addition of log vectors)
Vector Operations

- Vector addition surprisingly effective
  - But insensitive to word order
- (Multiplication is addition of log vectors)
- Tensors have compositional structure
  (Coecke et al., 2010; Baroni et al., 2013)
  - ... but high-order tensors are hard to learn
Neural Composition

- Neural net:
  - Sequence of vectors as input
  - One vector as output

Many architectures...
- LSTM
- Tree-LSTM
- DIORA (Deep Inside-Outside Recursive Autoencoder)
- RNNG (Recursive Neural Network Grammar)
...
Neural Composition

- Neural net:
  - Sequence of vectors as input
  - One vector as output

- Many architectures...
  - LSTM
  - Tree-LSTM
  - DIORA (Deep Inside-Outside Recursive Autoencoder)
  - RNNG (Recursive Neural Network Grammar)
  - ...

Distinct Meanings

- Every fluffy dog barked
- Every dog that barked is fluffy
Distinct Meanings

- Every fluffy dog barked
- Every dog that barked is fluffy
- The fluffy dog barked
- The dog that barked is fluffy
Distinct Meanings

- Every fluffy dog barked
- Every dog that barked is fluffy
- The fluffy dog barked
- The dog that barked is fluffy
- Every student passed the exam
- No student failed the exam
Combinatorics of Composition

- A {dog, cat, ...} {saw, chased, ...} a {dog, cat, ...}

Exponential growth in distinct meanings
Combinatorics of Composition

- A \{\text{dog, cat, ...}\} \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}

- A \{\text{dog, cat, ...}\} \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}
  which \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}
Combinatorics of Composition

- A \{\text{dog, cat, ...}\} \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}
- A \{\text{dog, cat, ...}\} \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}
  which \{\text{saw, chased, ...}\} a \{\text{dog, cat, ...}\}
- Exponential growth in distinct meanings
Limits of Vector Representations

- Exponential growth in distinct meanings

- Magnitudes grow exponentially
- Sensitive to arbitrarily small changes
- Lossy representations can still be useful, but don't give us a full semantic theory
Limits of Vector Representations

- Exponential growth in distinct meanings

- For vectors, one of the following must hold:
  - Magnitudes grows exponentially
  - Sensitive to arbitrarily small changes
  - Lossy

Lossy representations can still be useful, but don’t give us a full semantic theory.
Limits of Vector Representations

- Exponential growth in distinct meanings

- For vectors, one of the following must hold:
  - Magnitudes grows exponentially
  - Sensitive to arbitrarily small changes
  - Lossy

- Lossy representations can still be useful, but don’t give us a full semantic theory
Compositional Semantics

- Compositional by design:
  - Predicate logic with lambda calculus
  - Semantic graphs with graph rewriting

Bring compositional structure into distributional semantics?
Compositional Semantics

- Compositional by design:
  - Predicate logic with lambda calculus
  - Semantic graphs with graph rewriting

- Bring compositional structure into distributional semantics?
Truth-Conditional Semantics
Truth-Conditional Semantics
Truth-Conditional Functions
Truth-Conditional Functions
Truth-Conditional Functions
Truth-Conditional Functions
Truth-Conditional Functions
Summary of What’s New

- Pixie: entity representation
- Word meanings as functions: pixie → probability of truth
Situation Semantics

\[ \begin{align*}
X \\
\text{pepper}(x)
\end{align*} \]
Situation Semantics

\[ x \]

pepper(x)
vegetable(x)
animal(x)
dog(x)
cat(x)
Situation Semantics

\[
x \xleftarrow{\text{ARG1}} y \xrightarrow{\text{ARG2}} z
\]

dog(x)  animal(x)
chase(x)  pursue(x)
cat(x)  

cat(z)  animal(z)
chase(z)  pursue(z)  dog(z)
Probabilistic Situation Semantics

\[ \in X \]

- dog(X)
- animal(X)
- chase(X)
- pursue(X)
- cat(X)

- chase(Y)
- dog(Y)
- cat(Y)
- animal(Y)

- cat(Z)
- animal(Z)
- chase(Z)
- pursue(Z)
- dog(Z)
Probabilistic Situation Semantics

\[ X, Y, Z \in \mathcal{X} \]

\[ T_{r,X}, T_{r,Y}, T_{r,Z} \in \{\bot, T\} \]

\[ \nu \]

\[ \text{AR}_1 \quad \text{AR}_2 \]
Probabilistic Situation Semantics

- World model: \( P(x, y, z) \)
  (Joint distribution of pixie-valued random variables)

- Lexical model: \( P(t_r, x | x) \)
  (Conditional distribution of truth-valued random variables, given a pixie)
Distributional Semantics

- What should the model learn?
- How can the model learn it?
Distributional Semantics

- What should the model learn?
  - Probabilistic situation semantics
- How can the model learn it?
Distributional Semantics

- What should the model learn?
  - Probabilistic situation semantics
- How can the model learn it?
  - Probabilistic graphical model
  - Data: semantic dependency graphs
Every picture tells a story
∀ x ∀ y ∀ z picture(x) ⇒ \[ \text{story}(z) \] \tell(y) \arg1(y, x) \arg2(y, z)

Dependency Minimal Recursion Semantics

∀ picture tell ∃ story

∀x∃y∃z picture(x) ⇒ [story(z) ∧ tell(y) ∧ ARG1(y, x) ∧ ARG2(y, z)]
\[ \forall x \exists y \exists z \text{ picture}(x) \Rightarrow [\text{story}(z) \land \text{tell}(y) \\
\land \text{ARG1}(y, x) \land \text{ARG2}(y, z)] \]

Functional Distributional Semantics

\[ \exists X \in \mathcal{X} \]

\[ \{ \bot, \top \} \]

\[ T_{r,X}, T_{r,Y}, T_{r,Z} \in \mathcal{Y} \]

\[ \text{dog} \xleftarrow{\text{ARG1}} \text{chase} \xrightarrow{\text{ARG2}} \text{cat} \]
Functional Distributional Semantics

\[ T_r, X \in \mathcal{X} \]

\[ T_r, Y \in \{ \bot, T \} \]

\[ T_r, Z \in \mathcal{V} \]

\[ R_X \in \mathcal{V} \]

\[ R_Y \in \mathcal{V} \]

\[ R_Z \in \mathcal{V} \]
Functional Distributional Semantics

- Latent situation semantics
  - World model: $\mathbb{P}(x, y, z)$
  - Lexical model: $\mathbb{P}(t_{r,x} | x)$

- Observed DMRS graphs
  - Extended lexical model: $\mathbb{P}(r_x | x) \propto \mathbb{P}(t_{r,x} | x)$
    (For simplicity, probability of utterance assumed proportional to probability of truth)
World Model

- Restricted Boltzmann Machine (binary vectors)
World Model

- Restricted Boltzmann Machine (binary vectors)
- Fabiani 2021 MPhil project: prototype version using real-valued vectors and Gaussian distributions
World Model

- Restricted Boltzmann Machine (binary vectors)
World Model

- Restricted Boltzmann Machine (binary vectors)

\[ P(s) \propto \exp(-E(s)) \]
World Model

- Restricted Boltzmann Machine (binary vectors)

\[ \mathbb{P}(s) \propto \exp \left( \sum_{x \rightarrow y \text{ in } s} w_{ij}^{(L)} x_i y_j \right) \]
Lexical Model

- Feedforward networks

\[ P(t^{(r,x)} | x) = \sigma(v^{(r)}_i x_i) \]
Lexical Model

- Feedforward networks
- \[ P(t^{(r,X)} | x) = \sigma(v^{(r)}_i x_i) \]
- \[ P(r^{(X)} | x) \propto P(t^{(r,X)} | x) \]
Functional Distributional Semantics

\[ \forall \mathbf{v} \in \mathbf{V} \]

- pepper \((x)\)
- vegetable \((x)\)
- animal \((x)\)
- dog \((x)\)
- cat \((x)\)

\[ \mathbf{h}(\mathbf{v}) = \mathbf{h}(\mathbf{v}') \]

\[ \mathbf{R}_X \]
\[ \mathbf{R}_Y \]
\[ \mathbf{R}_Z \]
\[ \frac{\partial}{\partial \theta} \log P(g) = \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log P(s) \right] \\
+ \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log P(g \mid s) \right] \]
Gradient Descent

\[
\frac{\partial}{\partial \theta} \log \mathbb{P}(g) = \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log \mathbb{P}(s) \right] + \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log \mathbb{P}(g \mid s) \right]
\]
Gradient Descent

\[ \frac{\partial}{\partial \theta} \log P(g) = \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log P(s) \right] + \mathbb{E}_{s \mid g} \left[ \frac{\partial}{\partial \theta} \log P(g \mid s) \right] \]

- Latent variables necessary but inconvenient
Gradient Descent

\[ \frac{\partial}{\partial \theta} \log P(g) = \mathbb{E}_{s|g} \left[ \frac{\partial}{\partial \theta} \log P(s) \right] + \mathbb{E}_{s|g} \left[ \frac{\partial}{\partial \theta} \log P(g | s) \right] \]

- Latent variables necessary but inconvenient
- Need approximation (e.g. amortised variational inference: train neural net to do it)
Pixie Autoencoder (Emerson, 2020)

- Generative model (world model & lexical model)
- Inference network (approximate inference)
Pixie Autoencoder (Emerson, 2020)

- Generative model (world model & lexical model)
- Inference network (approximate inference)
- Unique selling point:
  - Truth-conditional distributional semantics
Training Needs Graphs

- Training needs dependency graphs, not raw text
Training Needs Graphs

- Training needs dependency graphs, not raw text
- WikiWoods
  - English Wikipedia, parsed into DMRS graphs
  - 31 million graphs (after preprocessing)
Training Needs Graphs

- Training needs dependency graphs, not raw text
- WikiWoods
  - English Wikipedia, parsed into DMRS graphs
  - 31 million graphs (after preprocessing)
  - (So far, only verbs with ARG1 & ARG2 nouns)
Lexical similarity: given two words (out of context), how similar are they?
Sanity Check: Lexical Similarity

- Lexical similarity: given two words (out of context), how similar are they?
- Competitive with state of the art
- Can distinguish similarity (mouse, rat) from relatedness (law, lawyer)
Similarity in Context (GS2011)

- Controlled semantic evaluation
- Starts to use expressiveness of functional model
<table>
<thead>
<tr>
<th>Student</th>
<th>Write</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>Spell</td>
<td>Name</td>
</tr>
<tr>
<td>Scholar</td>
<td>Write</td>
<td>Book</td>
</tr>
<tr>
<td>Scholar</td>
<td>Spell</td>
<td>Book</td>
</tr>
</tbody>
</table>
Pixie Autoencoder for GS2011

\[ P(t, Y | t, X, t, Y, t, Z) \]
Pixie Autoencoder for GS2011

\[ P(t_{a,Y} | t_{p,X}, t_{q,Y}, t_{r,Z}) \]
Pixie Autoencoder for GS2011

\[ P \left( t_{\text{spell}}, Y \mid t_{\text{student}}, X, t_{\text{write}}, Y, t_{\text{name}}, Z \right) \]
Pseudo-logical form: (employer provide training)

- “an employer provides training .”
- “employer provides training .”
- “an employer provides a training .”
- “a employer provides a training .”
- “employers provide training .”
- “employers provide trainings .”
- “training is provided by an employer .”
- “trainings are provided by employers .”
- ...
## GS2011 Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skip-gram (vector addition)</td>
<td>.348</td>
</tr>
<tr>
<td>BERT (with tuned template strings)</td>
<td>.446</td>
</tr>
<tr>
<td>Pixie Autoencoder</td>
<td>.504</td>
</tr>
</tbody>
</table>

- Smaller model, less data, better performance
RELPRON Dataset (Rimell et al., 2016)

- Controlled semantic evaluation
- Starts to use expressiveness of functional model
RELPRON Dataset (Rimell et al., 2016)

- Controlled semantic evaluation
- Starts to use expressiveness of functional model
- Large gap between human performance (~100%) and state of the art (~50%)
telescope: device that astronomers use
saw: device that cuts wood
philosopher: person that defends rationalism
survivor: person that helicopter saves
farming: activity that soil supports
...
RELPRON Dataset (Rimell et al., 2016)

telescope

device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
activity that soil supports
...

saw device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
activity that soil supports
...

RELPRON Dataset (Rimell et al., 2016)
RELPRON Dataset (Rimell et al., 2016)

- philosopher: device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
activity that soil supports
...

RELPRON Dataset (Rimell et al., 2016)

soil

device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
activity that soil supports

...
RELPRON Dataset (Rimell et al., 2016)

soil

device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
activity that soil supports
...

Logical Inference for RELPRON

X \rightarrow ARG1 \rightarrow Y \rightarrow ARG2 \rightarrow Z

Tp,X \rightarrow Tq,Y \rightarrow Tr,Z

philosopher, person, defend, rationalism
Logical Inference for RELPRON

\[ \mathbb{P}(t_{a,X} | t_{p,X}, t_{q,Y}, t_{r,Z}) \]
Logical Inference for RELPRON

\[
P(t_{a,X} \mid t_{p,X}, t_{q,Y}, t_{r,Z})
\]

\[
P(t_{philosopher,X} \mid t_{person,X}, t_{defend,Y}, t_{rationalism,Z})
\]
Pseudo-logical form: (person that defend rationalism)

- “A person that defends rationalism is a [MASK] .”
- “Person that defends rationalism is [MASK] .”
- “A person that defends a rationalism is a [MASK] .”
- “People that defend rationalisms are [MASK] .”
- “A [MASK] is a person that defends rationalism .”
- “A [MASK] is a person that defends a rationalism .”
- “A person that defends rationalism .”
- “A person that defends a rationalism .”
- …
## RELPRON Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simp. Prac. Lex. Func. (Rimell et al., 2016)</td>
<td>.497</td>
</tr>
<tr>
<td>Dependency vectors (Czarnowska et al., 2019)</td>
<td>.439</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>.474</td>
</tr>
<tr>
<td>BERT (with carefully tuned template strings)</td>
<td>.186</td>
</tr>
<tr>
<td>BERT &amp; Word2Vec ensemble</td>
<td>.479</td>
</tr>
<tr>
<td>Pixie Autoencoder</td>
<td>.189</td>
</tr>
<tr>
<td>Pixie Autoencoder &amp; Word2Vec ensemble</td>
<td>.489</td>
</tr>
</tbody>
</table>
## RELPRON Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simp. Prac. Lex. Func. (Rimell et al., 2016)</td>
<td>.497</td>
</tr>
<tr>
<td>Dependency vectors (Czarnowska et al., 2019)</td>
<td>.439</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>.474</td>
</tr>
<tr>
<td>BERT (with carefully tuned template strings)</td>
<td>.186</td>
</tr>
<tr>
<td>BERT &amp; Word2Vec ensemble</td>
<td>.479</td>
</tr>
<tr>
<td>Pixie Autoencoder</td>
<td>.189</td>
</tr>
<tr>
<td>Pixie Autoencoder &amp; Word2Vec ensemble</td>
<td>.489</td>
</tr>
</tbody>
</table>
## RELPRON Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simp. Prac. Lex. Func. (Rimell et al., 2016)</td>
<td>0.497</td>
</tr>
<tr>
<td>Dependency vectors (Czarnowska et al., 2019)</td>
<td>0.439</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>0.474</td>
</tr>
<tr>
<td>BERT (with carefully tuned template strings)</td>
<td>0.186</td>
</tr>
<tr>
<td>BERT &amp; Word2Vec ensemble</td>
<td>0.479</td>
</tr>
<tr>
<td>Pixie Autoencoder</td>
<td>0.189</td>
</tr>
<tr>
<td>Pixie Autoencoder &amp; Word2Vec ensemble</td>
<td>0.489</td>
</tr>
</tbody>
</table>
Pixie Autoencoder compared to BERT:
- More data efficient (1.2% no. tokens)
- Doesn’t require tuning to apply
- More “different” from Word2Vec
Pixie Autoencoder compared to BERT:
- More data efficient (1.2% no. tokens)
- Doesn’t require tuning to apply
- More “different” from Word2Vec

Word2Vec still state of the art
- Error analysis: good at relatedness
- Need “topic” in world model?
Ongoing/Future Work

- Joint learning with grounded data
- Joint learning with lexical resources
- More efficient model (continuous pixies)
- Latent variable for “topic”
- Correlated truth values (for pragmatics)
- Deeper networks (for polysemy)
- Semi-compositional idioms
- More general logical inferences
Ongoing/Future Work

- Joint learning with grounded data
- Joint learning with lexical resources
- More efficient model (continuous pixies)
- Latent variable for “topic”
- Correlated truth values (for pragmatics)
- Deeper networks (for polysemy)
- Semi-compositional idioms
- More general logical inferences
Joint Learning with Grounded Data

- Fundamental distinction between words and entities
Joint Learning with Grounded Data

- Fundamental distinction between words and entities
- Vector space models:
  - Early fusion, late fusion, cross-modal maps...
Joint Learning with Grounded Data

- Fundamental distinction between words and entities
- Vector space models:
  - Early fusion, late fusion, cross-modal maps...
- Functional Distributional Semantics:
  - Text → pixies are latent
  - Grounded data → pixies are observed
Visual Genome Dataset

“couple cutting cake”
Visual Genome Dataset

“couple cutting cake”
Visual Genome Semantics

- Liu 2021 MPhil project: learn functional model from Visual Genome
- (Not joint learning... yet)
Liu 2021 MPhil project: learn functional model from Visual Genome

(Not joint learning... yet)

<table>
<thead>
<tr>
<th>Model</th>
<th>MEN</th>
<th>SL999</th>
<th>GS2011</th>
<th>RELPRON</th>
</tr>
</thead>
<tbody>
<tr>
<td>VG-count (Herbelot, 2020)</td>
<td>.336</td>
<td>.224</td>
<td>.063</td>
<td>.038</td>
</tr>
<tr>
<td>EVA (Herbelot, 2020)</td>
<td>.543</td>
<td>.390</td>
<td>.068</td>
<td>.032</td>
</tr>
<tr>
<td>Functional</td>
<td>.639</td>
<td>.431</td>
<td>.171</td>
<td>.117</td>
</tr>
</tbody>
</table>
Conclusion

- Distributional semantics: more than just similarity
- Compositionality ≠ Context dependence
- Vectors: useful but have fundamental limitations
- Truth-conditional distributional semantics: feasible, but long road ahead!