Distributional Semantics and Vector Codes for Concepts and their Combinations

Alessandro Lenci

COmputational LINGuistics Laboratory
University of Pisa
Dipartimento di Filologia, Letteratura e Linguistica (FiLeLi)

Concepts and Experience

- Concepts and their features are derived from our experience
  - sensory-motor and affective experiences (exposure to objects and events)

  - linguistic experiences (exposure to linguistic input)

    ... so we went outside, picked several red cherries and ate them ...
    ... the colour of an orange pink sunset and an indulgent length of rich, red cherry fruit with hints of almonds on the dry finish...
    ... place the muffin and cherries at the bottom of a bowl, add the ice cream ...
    ... topped with sweet and sticky black cherries on a smooth chocolate sauce ...

Concepts and Experience

- Concepts and their features are derived from our experience
  - sensory-motor and affective experiences (exposure to objects and events)

  - linguistic experiences (exposure to linguistic input)

. . . so we went outside, picked several red cherries and ate them . . .
. . . the colour of an orange pink sunset and an indulgent length of rich, red cherry fruit with hints of almonds on the dry finish. . .
. . . place the muffin and cherries at the bottom of a bowl, add the ice cream . . .
. . . topped with sweet and sticky black cherries on a smooth chocolate sauce . . .

Linguistic Experience and Meaning

- Both embodied and linguistic information contributes to conceptual representations (Barsalou et al. 2008, Borghi & Cimatti 2009, Borghi & Binkofsky 2014, Vigliocco et al. 2009)
  - representational pluralism (Dove 2009, 2014, Scorolli et al. 2011)
- Linguistic experience is more salient for certain areas of semantic memory
  - abstract terms
    - *justice*, *tax*, etc.
  - concrete terms for which we have no direct acquaintance with theirreferents
    - *aardvark*, *cyclotron*, etc.
  - verb manings (cf. Syntactic Bootstrapping; Landau & Gleitman (1985); McDonald & Ramscar (2001); Fisher & Gleitman 2002)
    - *The man gorped Mary the book*
    - *John gorped that he was unhappy*

Linguistic Experience and Meaning

- Both embodied and linguistic information contributes to conceptual representations (Barsalou et al. 2008, Borghi & Cimatti 2009, Borghi & Binkofsky 2014, Vigliocco et al. 2009)
  - representational pluralism (Dove 2009, 2014, Scorolli et al. 2011)
- Linguistic experience is more salient for certain areas of semantic memory
  - abstract terms
    - *justice*, *tax*, etc.
  - concrete terms for which we have no direct acquaintance with theirreferents
    - *aardvark*, *cyclotron*, etc.
  - verb manings (cf. Syntactic Bootstrapping; Landau & Gleitman (1985); McDonald & Ramscar (2001); Fisher & Gleitman 2002)
    - *The man gorped Mary the book*
    - *John gorped that he was unhappy*

Representing Concepts with Symbols

- Concepts are traditionally represented with structures of formal symbols

dog = [+ANIMATE, -ARTIFACT, +BARK,+FOUR_LEGS, ...]

$enter = [_{\text{EVENT}}\ \text{GO}\ ([\text{THING}_i], [_{\text{PATH}}\ \text{TO}\ ([_{\text{PLACE}}\ \text{IN}\ ([\text{THING}_j])])])]$

$$book = \begin{bmatrix} \text{ARGSTR} & = & \begin{bmatrix} \text{ARG1} = \textbf{x:info} \\ \text{ARG2} = \textbf{y:physobj} \end{bmatrix} \\ \text{QUALIA} & = & \begin{bmatrix} \textbf{info}\cdot\textbf{physobj\_lcp} \\ \text{FORMAL} = \textbf{hold(y,x)} \\ \text{CONST} = \textbf{part\_of(z:page,y)} \\ \text{TELIC} = \textbf{read(e,w,x)} \\ \text{AGENT} = \textbf{write(e,v,x)} \end{bmatrix} \end{bmatrix}$$

Representing Concepts with Symbols

- Concepts are traditionally represented with structures of formal symbols

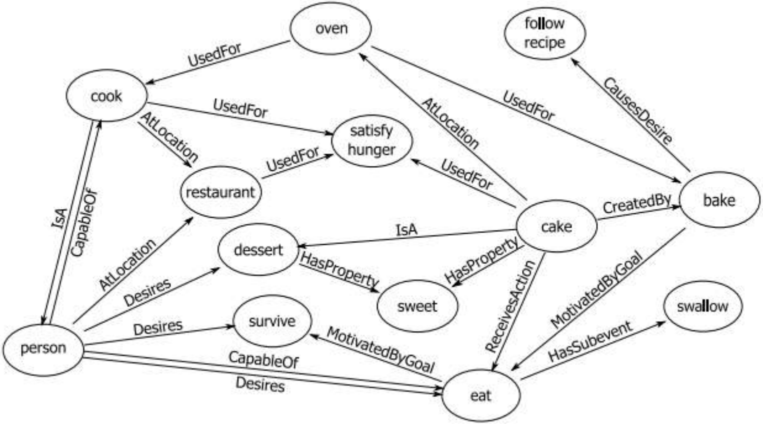
dog = [+ANIMATE, -ARTIFACT, +BARK,+FOUR_LEGS, ...]

$enter = [_{\text{EVENT}}\ \text{GO}\ ([\text{THING}_i], [_{\text{PATH}}\ \text{TO}\ ([_{\text{PLACE}}\ \text{IN}\ ([\text{THING}_j])])])]$

$$book = \begin{bmatrix} \text{ARGSTR} & = & \begin{bmatrix} \text{ARG1} = \textbf{x:info} \\ \text{ARG2} = \textbf{y:physobj} \end{bmatrix} \\ \text{QUALIA} & = & \begin{bmatrix} \textbf{info}\cdot\textbf{physobj\_lcp} \\ \text{FORMAL} = \textbf{hold(y,x)} \\ \text{CONST} = \textbf{part\_of(z:page,y)} \\ \text{TELIC} = \textbf{read(e,w,x)} \\ \text{AGENT} = \textbf{write(e,v,x)} \end{bmatrix} \end{bmatrix}$$

Representing Concepts with Symbols

- Concepts are traditionally represented with structures of formal symbols



The Limits of Symbolic Representations

- Discrete, qualitative and rigid
  - hard to tackle gradience, fuzziness, etc.
- Too stipulative and a priori
  - hard to identify the primitives and the atomic elements of such representations (e.g., the repertoire of semantic types)
- Limited explicative power of empirical linguistic and cognitive facts
  - many semantic phenomena hard to tackle (e.g. context meaning shifts, etc.) or require to complicate the semantic machinery
- Inherently amodal
  - hard to model the integration of multimodal sources of information
- Lack of sound methods to learn them
  - semantic acquisition (often) left out of the picture

The Limits of Symbolic Representations

- Discrete, qualitative and rigid
  - hard to tackle gradience, fuzziness, etc.
- Too stipulative and a priori
  - hard to identify the primitives and the atomic elements of such representations (e.g., the repertoire of semantic types)
- Limited explicative power of empirical linguistic and cognitive facts
  - many semantic phenomena hard to tackle (e.g. context meaning shifts, etc.) or require to complicate the semantic machinery
- Inherently amodal
  - hard to model the integration of multimodal sources of information
- Lack of sound methods to learn them
  - semantic acquisition (often) left out of the picture

The Limits of Symbolic Representations

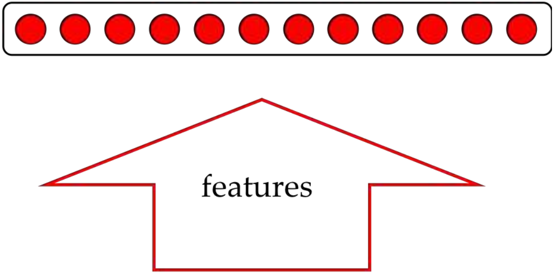
- Discrete, qualitative and rigid
  - hard to tackle gradience, fuzziness, etc.
- Too stipulative and a priori
  - hard to identify the primitives and the atomic elements of such representations (e.g., the repertoire of semantic types)
- Limited explicative power of empirical linguistic and cognitive facts
  - many semantic phenomena hard to tackle (e.g. context meaning shifts, etc.) or require to complicate the semantic machinery
- Inherently amodal
  - hard to model the integration of multimodal sources of information
- Lack of sound methods to learn them
  - semantic acquisition (often) left out of the picture

The Limits of Symbolic Representations

- Discrete, qualitative and rigid
  - hard to tackle gradience, fuzziness, etc.
- Too stipulative and a priori
  - hard to identify the primitives and the atomic elements of such representations (e.g., the repertoire of semantic types)
- Limited explicative power of empirical linguistic and cognitive facts
  - many semantic phenomena hard to tackle (e.g. context meaning shifts, etc.) or require to complicate the semantic machinery
- Inherently amodal
  - hard to model the integration of multimodal sources of information
- Lack of sound methods to learn them
  - semantic acquisition (often) left out of the picture

Representing Concepts with Vectors

- Concepts are represented with real-valued vectors



… so we went outside, picked several red cherries and ate them … the colour of an orange pink sunset and an indulgent length of rich, red cherry fruit with hints of almonds on the dry finish …

Representing Concepts with Vectors

- Continuous and distributed representations
  - easier to tackle gradience, fuzziness, similarity, analogy-based processes, etc.
- Less stipulative
  - no need to assume a priori semantic primitives or features
- Availability of methods to learn semantic representations
  - semantic acquisition back in the game
- Not necessarily amodal
  - easy integration of multimodal features

BUT…

still limited explicative power of several semantic facts

Representing Concepts with Vectors

- Continuous and distributed representations
  - easier to tackle gradience, fuzziness, similarity, analogy-based processes, etc.
- Less stipulative
  - no need to assume a priori semantic primitives or features
- Availability of methods to learn semantic representations
  - semantic acquisition back in the game
- Not necessarily amodal
  - easy integration of multimodal features

BUT…

still limited explicative power of several semantic facts

Representing Concepts with Vectors

- Continuous and distributed representations
  - easier to tackle gradience, fuzziness, similarity, analogy-based processes, etc.
- Less stipulative
  - no need to assume a priori semantic primitives or features
- Availability of methods to learn semantic representations
  - semantic acquisition back in the game
- Not necessarily amodal
  - easy integration of multimodal features

BUT...

still limited explicative power of several semantic facts

Representing Concepts with Vectors

- Continuous and distributed representations
  - easier to tackle gradience, fuzziness, similarity, analogy-based processes, etc.
- Less stipulative
  - no need to assume a priori semantic primitives or features
- Availability of methods to learn semantic representations
  - semantic acquisition back in the game
- Not necessarily amodal
  - easy integration of multimodal features

BUT...

still limited explicative power of several semantic facts

Representing Concepts with Vectors

- Continuous and distributed representations
  - easier to tackle gradience, fuzziness, similarity, analogy-based processes, etc.
- Less stipulative
  - no need to assume a priori semantic primitives or features
- Availability of methods to learn semantic representations
  - semantic acquisition back in the game
- Not necessarily amodal
  - easy integration of multimodal features

BUT...

still limited explicative power of several semantic facts

What is Distributional Semantics?

Lenci (2018), "Distributional Models of Word Meaning", *Annual Review of Linguistics*, 4

Distributional Semantics

The study of how distributional information, that is the statistical distribution of lexemes in linguistic contexts, can be used to model semantic facts

Distributional Representation

The distributional representation of a lexical item is an n-dimensional distributional vector, whose components represent its co-occurrences with linguistic contexts

The Distributional Hypothesis (DH)

Lexemes with similar distributional properties have similar meanings

What is Distributional Semantics?

Lenci (2018), “Distributional Models of Word Meaning”, *Annual Review of Linguistics*, 4

Distributional Semantics

The study of how distributional information, that is the statistical distribution of lexemes in linguistic contexts, can be used to model semantic facts

Distributional Representation

The distributional representation of a lexical item is an *n*-dimensional distributional vector, whose components represent its co-occurrences with linguistic contexts

The Distributional Hypothesis (DH)

Lexemes with similar distributional properties have similar meanings

What is Distributional Semantics?

Lenci (2018), "Distributional Models of Word Meaning", *Annual Review of Linguistics*, 4

Distributional Semantics

The study of how distributional information, that is the statistical distribution of lexemes in linguistic contexts, can be used to model semantic facts

Distributional Representation

The distributional representation of a lexical item is an *n*-dimensional distributional vector, whose components represent its co-occurrences with linguistic contexts

The Distributional Hypothesis (DH)

Lexemes with similar distributional properties have similar meanings

P. Garvin, (1962), “Computer participation in linguistic research”, *Language*, 38(4): 385-389

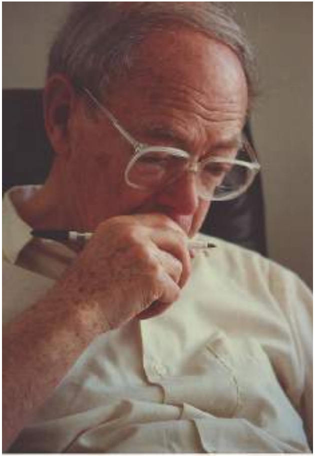
Distributional semantics is predicated on the assumption that linguistic units with certain semantic similarities also share certain similarities in the relevant environments.

If therefore relevant environments can be previously specified, it may be possible to group automatically all those linguistic units which occur in similarly definable environments, and it is assumed that these automatically produced groupings will be of semantic interest.

The Pioneers of Distributional Semantics

Distributionalism in linguistics

Zellig S. Harris



If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference in meaning correlates with difference of distribution.

(Harris 1954: 156)

The Pioneers of Distributional Semantics

Distributionalism in linguistics

John R. Firth

As Wittgenstein says, 'the meaning of words lies in their use.' The day to day practice of playing language games recognizes customs and rules. It follows that a text in such established usage may contain sentences such as 'Don't be such an ass!', 'you silly ass!', 'What an ass he is!' In these examples, the word ass is in familiar and habitual company, commonly collocated with you silly –, he is a silly –, don't be such an –. You shall know a word by the company it keeps!

(Firth 1957: 11)

The Pioneers of Distributional Semantics

Distributionalism in cognitive science

George A. Miller

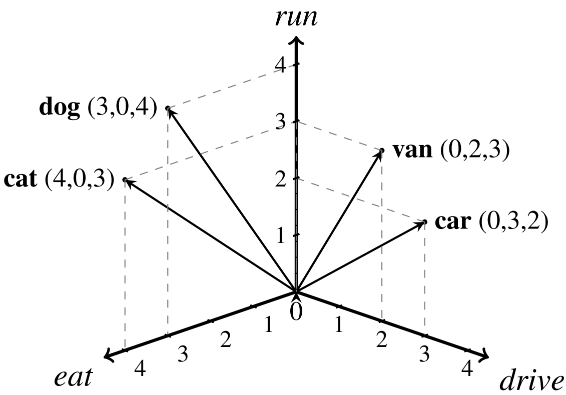
The contextual representation of a word is knowledge of how that word is used. [...] That is to say, a word's contextual representation [...] is an abstract cognitive structure that accumulates from encounters with the word in various (linguistic) contexts. [...] Two words are semantically similar to the extent that their contextual representations are similar.

(Miller and Charles 1991: 5)

From Linguistic contexts . . .

. . . dig a [*hole. The*	**car**	*drove away*] leaving behind . . .
. . . to directly [*drive the*	**car**	*wheel angle*] 3. Force . . .
. . . to pet [*the family's*	**cat**	*and dog,*] who tended . . .
. . . and then [*wanted a*	**cat**	*to eat*] the many . . .
. . . bank, children [*playing with*	**dogs**	*and a*] man leading. . . .
. . . vegetable material [*and enzymes.*	**Dogs**	*also eat*] fruit, berries . . .
. . . hubby once [*ate the*	**dog**	*food and*] asked for . . .
. . . go down [*as the*	**van**	*drove off.*] As he . . .
. . . heavy objects, [*driving transit*	**vans**	*, wiring plugs*] and talking . . .
. . . of the [*fast food*	**van**	*being located*] outside their . . .
. . . each of [*the six*	**van**	*wheels , and*] also under . . .

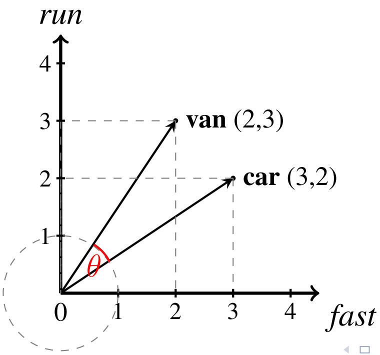
. . . to Distributional Vectors



Distributional Similarity

- The distributional similarity between two lexemes *u* and *v* is measured with the similarity between their distributional vectors **u** and **v**

Cosine $$\frac{\mathbf{x} \cdot \mathbf{y}}{|\,\mathbf{x}\,||\,\mathbf{y}\,|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$



Distributional Similarity

- Given the Distributional Hypothesis, semantically similar lexemes are expected to be distributionally similar

car	1			
cat	0.33	1		
dog	0.44	0.96	1	
van	0.92	0.50	0.66	1
	car	*cat*	*dog*	*van*

Distributional Semantic Models

Distributional Semantic Models (DSMs)

Computational methods to learn distributional representations from text corpora

Model name	Description
Latent Semantic Analysis (LSA)[a]	word-by-document matrix reduced with SVD
Hyperspace Analogue of Language (HAL)[b]	window-based model with directed collocates
Distributional Memory (DM)[c]	tensor model with dependency-typed collocates
Topic Models[d]	word-by-document matrix reduced with Bayesian inference
Random Indexing[e]	accumulation of contexts encoded with random vectors
word2vec (CBOW, skipgram)[f]	neural network model predicting neighboring words
Global Vectors (GloVe)[g]	word-by-word matrix reduced with weighted regression

[a] Landauer & Dumais (1997); [b] Lund & Burgess (1996); [c] Baroni & Lenci (2010); [d] Griffiths et al. (2007); [e] Kanerva et al. (2000); [f] Mikolov et al. (2013a,b); [g] Pennington et al. (2014).

Distributional Semantic Models

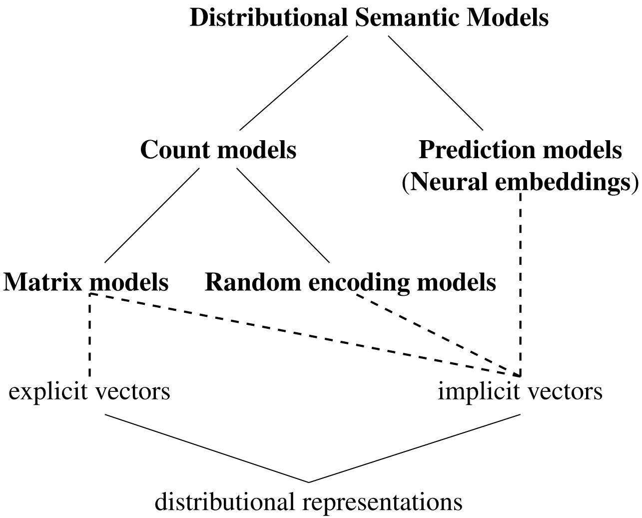
Distributional Semantic Models (DSMs)

Computational methods to learn distributional representations from text corpora

Model name	**Description**
Latent Semantic Analysis (LSA)[a]	word-by-document matrix reduced with SVD
Hyperspace Analogue of Language (HAL)[b]	window-based model with directed collocates
Distributional Memory (DM)[c]	tensor model with dependency-typed collocates
Topic Models[d]	word-by-document matrix reduced with Bayesian inference
Random Indexing[e]	accumulation of contexts encoded with random vectors
word2vec (CBOW, skipgram)[f]	neural network model predicting neighboring words
Global Vectors (GloVe)[g]	word-by-word matrix reduced with weighted regression

[a] Landauer & Dumais (1997); [b] Lund & Burgess (1996); [c] Baroni & Lenci (2010); [d] Griffiths et al. (2007); [e] Kanerva et al. (2000); [f] Mikolov et al. (2013a,b); [g] Pennington et al. (2014).

DSMs Classified by Learning Method



Explicit vs. Implicit Vectors

Explicit vectors

High-dimensional, sparse vectors in which each dimension corresponds to a distinct linguistic context

$$\begin{array}{r c}
 & \begin{array}{ccccccccc} \textit{bite} & \textit{buy} & \textit{drive} & \textit{eat} & \textit{get} & \textit{live} & \textit{park} & \textit{ride} & \textit{tell} \end{array} \\
\begin{array}{r} \textit{bike} \\ \textit{car} \\ \textit{dog} \\ \textit{lion} \end{array} & \begin{pmatrix} 0 & 9 & 0 & 0 & 0 & 0 & 8 & 6 & 0 \\ 0 & 0 & 8 & 0 & 15 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 & 7 & 0 & 0 & 1 \\ 6 & 0 & 0 & 1 & 0 & 3 & 0 & 0 & 0 \end{pmatrix}
\end{array}$$

Explicit vs. Implicit Vectors

Implicit vectors (aka embeddings)

Low-dimensional, dense vectors of latent dimensions

$$\begin{array}{l} \textit{bike} \\ \textit{car} \\ \textit{dog} \\ \textit{lion} \end{array} \begin{pmatrix} -0.57 & 0.24 & -0.78 & -0.06 \\ -0.72 & 0.31 & 0.62 & -0.05 \\ -0.32 & -0.83 & 0.01 & -0.45 \\ -0.23 & -0.39 & -0.01 & 0.89 \end{pmatrix}$$

- Vector dimensions do not have a direct interpretation
- Interpretability comes from relations between vectors in semantic space

Distributional Semantics and Semantic Similarity

- The Distributional Hypothesis is couched in terms of similarity, but DSMs are actually more biased towards the much vaguer notion of semantic relatedness

Target	Neighbors[a]
car	*truck, vehicle, driving, garage, drive, jeep, windshield, driver, drove, bike*
smart	*dumb, clever, stupid, intelligent, pretty, enough, tough, you, think, cute*
eat	*hungry, eating, ate, eaten, eats, food, meal, starving, lunch, delicious*

[a] Nearest neighbors in CBOW ordered from left to right by similarity.

Cf. http://meshugga.ugent.be/snaut-english/ (Mandera et al. 2017).

Distributional Semantics and Semantic Similarity

- The Distributional Hypothesis is couched in terms of similarity, but DSMs are actually more biased towards the much vaguer notion of semantic relatedness

Target	**Neighbors**[a]
car	*truck*, *vehicle*, *driving*, *garage*, *drive*, *jeep*, *windshield*, *driver*, *drove*, *bike*
smart	*dumb*, *clever*, *stupid*, *intelligent*, *pretty*, *enough*, *tough*, *you*, *think*, *cute*
eat	*hungry*, *eating*, *ate*, *eaten*, *eats*, *food*, *meal*, *starving*, *lunch*, *delicious*

[a] Nearest neighbors in CBOW ordered from left to right by similarity.

Cf. http://meshugga.ugent.be/snaut-english/ (Mandera et al. 2017).

Semantic Similarity and Cognitive Modeling

- Distributional representations are successfully used to model behavioral data in psycholinguistic and neurocognitive lexical tasks involving semantic relatedness (Mandera et al. 2017)
    - association norms (Andrews et al. 2009, Mandera et al. 2017)
    - noun categorization (Baroni & Lenci 2010, Riordan & Jones 2011)
    - semantic priming (Jones et al. 2006, Mandera et al. 2017)
    - fMRI activations (Mitchell et al. 2008, Anderson et al. 2017)

The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only . . .)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only . . .)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only ...)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only . . .)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

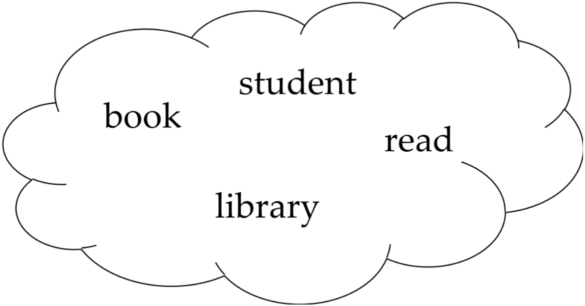
The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only . . .)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

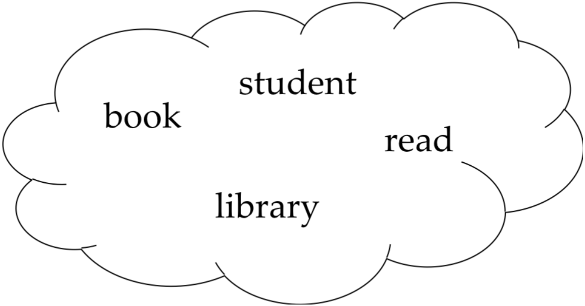
The Main Characters of Distributional Semantics

- Distributional semantics offers both a model to represent meaning with vectors and computational methods to learn such representations from language data (but not only ...)
  - cf. Multimodal Distributional Semantics (Feng & Lapata 2012, Bruni et al. 2014)
- Distributional representations are continuous and gradable
- Distributional semantics is based on a contextual and usage-based view of meaning
- The output of DSMs is a measure of semantic similarity/relatedness
- Distributional semantics is primarily a model of the lexicon

From Words . . .



...to Sentences ...

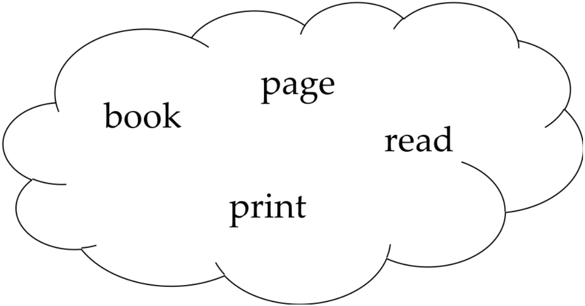


A student reads a book in a library.

Memories of Events

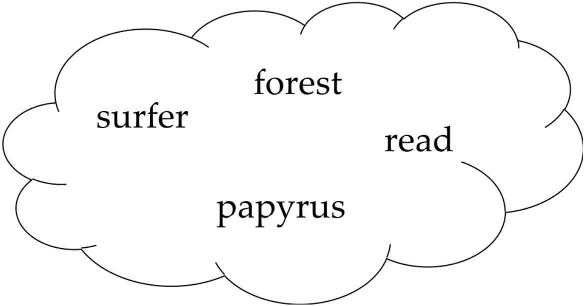
“Language is not merely a bag of words”

(Harris 1954: 156)

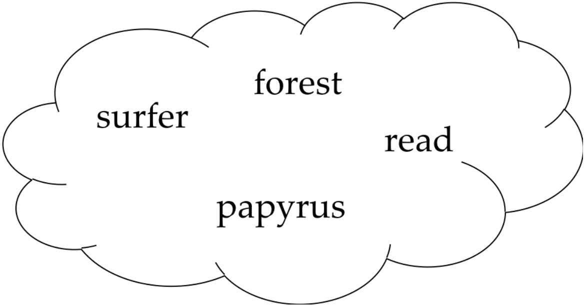


?????????

New Sentences, New Events

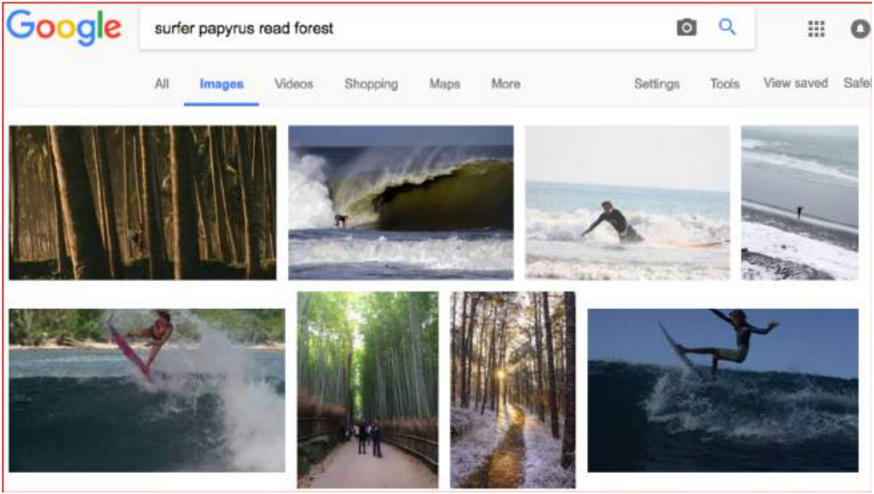


New Sentences, New Events



A surfer reads a papyrus in a forest.

New Sentences, new Memories



Language Comprehension (I)

- The brain is able to combine concepts to form coherent semantic representations of situations and events (semantic binding)
- Syntactic structure is a powerful tool to allow such a combinatorial capacity, but it is not strictly (always) necessary (provided proper background, pragmatic knowledge is available)
  - Humphries, C., Binder, J. R., et al. (2006). "Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension", *Journal of Cognitive Neuroscience*, 18(4), 665-679
  - Jackendoff, R. & Wittenberg, E. (2014). "What You Can Say Without Syntax: A Hierarchy of Grammatical Complexity". In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring Linguistic Complexity* (pp. 65–82), Oxford University Press
- Concepts have a combinatorial structure that allows them to be bound together and form coherent complex representations
  - combinatorial semantic constraints are linked to, but independent from syntactic ones (cf. Jackendoff 1997, 2002; Hagoort 2013)

Language Comprehension (I)

- The brain is able to combine concepts to form coherent semantic representations of situations and events (semantic binding)
- Syntactic structure is a powerful tool to allow such a combinatorial capacity, but it is not strictly (always) necessary (provided proper background, pragmatic knowledge is available)
  - Humphries, C., Binder, J. R., et al. (2006). "Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension", *Journal of Cognitive Neuroscience*, 18(4), 665-679
  - Jackendoff, R. & Wittenberg, E. (2014). "What You Can Say Without Syntax: A Hierarchy of Grammatical Complexity". In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring Linguistic Complexity* (pp. 65–82), Oxford University Press
- Concepts have a combinatorial structure that allows them to be bound together and form coherent complex representations
  - combinatorial semantic constraints are linked to, but independent from syntactic ones (cf. Jackendoff 1997, 2002; Hagoort 2013)

Language Comprehension (I)

- The brain is able to combine concepts to form coherent semantic representations of situations and events (semantic binding)
- Syntactic structure is a powerful tool to allow such a combinatorial capacity, but it is not strictly (always) necessary (provided proper background, pragmatic knowledge is available)
  - Humphries, C., Binder, J. R., et al. (2006). “Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension”, *Journal of Cognitive Neuroscience*, 18(4), 665-679
  - Jackendoff, R. & Wittenberg, E. (2014). “What You Can Say Without Syntax: A Hierarchy of Grammatical Complexity”. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring Linguistic Complexity* (pp. 65–82), Oxford University Press
- Concepts have a combinatorial structure that allows them to be bound together and form coherent complex representations
  - combinatorial semantic constraints are linked to, but independent from syntactic ones (cf. Jackendoff 1997, 2002; Hagoort 2013)

Gluing Symbols Together

- In symbolic representations, semantic composition is modeled with function-argument structures
  - "Every theory of semantics back to Frege acknowledges that word meanings may contain variables that are satisfied by arguments expressed elsewhere in the sentence" (Jackendoff 2002: 360)
- The output of the semantic composition for sentence can be a truth value, a proposition or an event

$$\begin{array}{lll} \textit{read} & \Rightarrow & [_{EVENT}\text{READ}([_{OBJECT}\text{X} : \text{Animate}], [_{OBJECT}\text{Y})] \\ \textit{student} & \Rightarrow & [_{OBJECT}\text{STUDENT}] \\ \textit{book} & \Rightarrow & [_{OBJECT}\text{BOOK}] \end{array}$$

$$\textit{A student reads a book} \Rightarrow [_{EVENT}\text{READ}([_{OBJECT}\text{STUDENT}], [_{OBJECT}\text{BOOK})]$$

Gluing Vectors together

- Lexical vectors are bound together and projected to phrase vectors with linear-algebraic operations (Mitchell and Lapata 2010, Baroni et al. 2014)
  - vector addition (Landauer & Dumais 1997)
  - tensor product (Smolensky 1990) and circular convolution (Jones & Mewhort 2007)
- Functional elements are represented with higher-order tensors and function-argument application with tensor by vector multiplication (Coecke et al. 2011, Baroni et al. 2014, Grefenstette & Sadrzadeh 2015, Rimell et al. 2016)

Gluing Vectors together

- Lexical vectors are bound together and projected to phrase vectors with linear-algebraic operations (Mitchell and Lapata 2010, Baroni et al. 2014)
  - vector addition (Landauer & Dumais 1997)
  - tensor product (Smolensky 1990) and circular convolution (Jones & Mewhort 2007)
- Functional elements are represented with higher-order tensors and function-argument application with tensor by vector multiplication (Coecke et al. 2011, Baroni et al. 2014, Grefenstette & Sadrzadeh 2015, Rimell et al. 2016)

Gluing Vectors together

- Compositionality as vector addition can not be the whole story...
  - *a cat chases a mouse* $\neq$ *a mouse chases a cat*
  - $\overrightarrow{a} + \overrightarrow{cat} + \overrightarrow{chases} + \overrightarrow{a} + \overrightarrow{mouse} = \overrightarrow{a} + \overrightarrow{mouse} + \overrightarrow{chases} + \overrightarrow{a} + \overrightarrow{cat}$
- ... still simple addition is generally the best performing model

General issue

We have solid intuitions about a word distributional representation and what it is useful for…

- i.e., a vector encoding its co-occurrences with other words that can be used to measure word similarity

…but it is not clear what the distributional representation of a sentence or discourse is and how to use it (besides measuring sentence similarity)

Gluing Vectors together

- Compositionality as vector addition can not be the whole story...
  - *a cat chases a mouse* $\neq$ *a mouse chases a cat*
  - $\overrightarrow{a} + \overrightarrow{cat} + \overrightarrow{chases} + \overrightarrow{a} + \overrightarrow{mouse} = \overrightarrow{a} + \overrightarrow{mouse} + \overrightarrow{chases} + \overrightarrow{a} + \overrightarrow{cat}$
- ... still simple addition is generally the best performing model

General issue

We have solid intuitions about a word distributional representation and what it is useful for…

- i.e., a vector encoding its co-occurrences with other words that can be used to measure word similarity

…but it is not clear what the distributional representation of a sentence or discourse is and how to use it (besides measuring sentence similarity)

Gluing Vectors together

- Compositionality as vector addition can not be the whole story...
  - *a cat chases a mouse* $\neq$ *a mouse chases a cat*
  - $\vec{a} + \overrightarrow{cat} + \overrightarrow{chases} + \vec{a} + \overrightarrow{mouse} = \vec{a} + \overrightarrow{mouse} + \overrightarrow{chases} + \vec{a} + \overrightarrow{cat}$
- ... still simple addition is generally the best performing model

General issue

We have solid intuitions about a word distributional representation and what it is useful for...

- i.e., a vector encoding its co-occurrences with other words that can be used to measure word similarity

...but it is not clear what the distributional representation of a sentence or discourse is and how to use it (besides measuring sentence similarity)

Language Comprehension (II)

- The comprehension of a sentence is an incremental process driven by the goal of constructing a coherent semantic representation of the event the speaker intends to communicate
- Sentences are partial descriptions of events
  - several details of events are left unspecified by the sentences describing them
  - implicit aspects can be (probabilistically) recovered or inferred thanks to our general knowledge about events and situations

    John surfed yesterday

    - John used a board
    - John was in the ocean or the sea
    - John wore a swimsuit or a wetsuit
    - ...

Understanding is predicting

Understanding a sentence allows us to make predictions

Language Comprehension (II)

- The comprehension of a sentence is an incremental process driven by the goal of constructing a coherent semantic representation of the event the speaker intends to communicate
- Sentences are partial descriptions of events
  - several details of events are left unspecified by the sentences describing them
  - implicit aspects can be (probabilistically) recovered or inferred thanks to our general knowledge about events and situations

John surfed yesterday

- John used a board
- John was in the ocean or the sea
- John wore a swimsuit or a wetsuit
- ...

Understanding is predicting

Understanding a sentence allows us to make predictions

Language Comprehension (II)

- The comprehension of a sentence is an incremental process driven by the goal of constructing a coherent semantic representation of the event the speaker intends to communicate
- Sentences are partial descriptions of events
  - several details of events are left unspecified by the sentences describing them
  - implicit aspects can be (probabilistically) recovered or inferred thanks to our general knowledge about events and situations

    John surfed yesterday

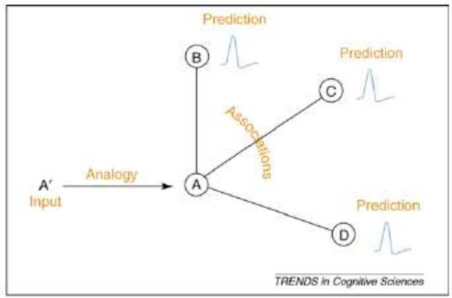
    - John used a board
    - John was in the ocean or the sea
    - John wore a swimsuit or a wetsuit
    - ...

Understanding is predicting

Understanding a sentence allows us to make predictions

Prediction in Cognition

Bar, M. (2007), "The proactive brain: using analogies and associations to generate predictions", *Trends in Cognitive Sciences*, 11(7), 280–289

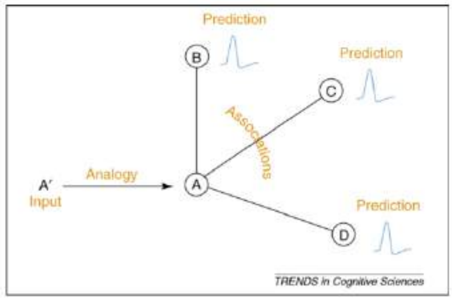


- The brain is constantly engaged in making predictions to anticipate events
- Predictions are memory-based, and rely on our previous experience about statistical associations between events and entities
- Predictions are carried out by detecting similarities between new inputs and stored associations

cf. Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". *Behavioral and Brain Sciences*, 36(3), 1–73

Prediction in Cognition

Bar, M. (2007), “The proactive brain: using analogies and associations to generate predictions”, *Trends in Cognitive Sciences*, 11(7), 280–289

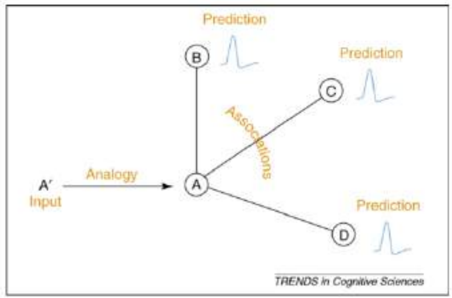


- The brain is constantly engaged in making predictions to anticipate events
- Predictions are memory-based, and rely on our previous experience about statistical associations between events and entities
- Predictions are carried out by detecting similarities between new inputs and stored associations

cf. Clark, A. (2013). “Whatever next? Predictive brains, situated agents, and the future of cognitive science”. *Behavioral and Brain Sciences*, 36(3), 1–73

Prediction in Cognition

Bar, M. (2007), "The proactive brain: using analogies and associations to generate predictions", *Trends in Cognitive Sciences*, 11(7), 280–289

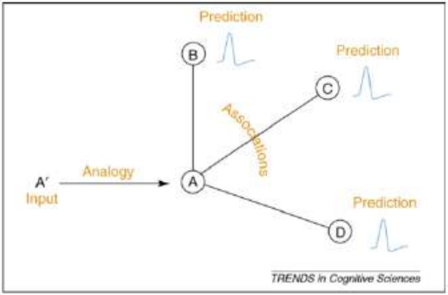


- The brain is constantly engaged in making predictions to anticipate events
- Predictions are memory-based, and rely on our previous experience about statistical associations between events and entities
- Predictions are carried out by detecting similarities between new inputs and stored associations

cf. Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". *Behavioral and Brain Sciences*, 36(3), 1–73

Prediction in Cognition

Bar, M. (2007), "The proactive brain: using analogies and associations to generate predictions", *Trends in Cognitive Sciences*, 11(7), 280–289



- The brain is constantly engaged in making predictions to anticipate events
- Predictions are memory-based, and rely on our previous experience about statistical associations between events and entities
- Predictions are carried out by detecting similarities between new inputs and stored associations

cf. Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". *Behavioral and Brain Sciences*, 36(3), 1–73

Language Comprehension (II)

(1) a. A student reads a book in a library.
b. A surfer reads a papyrus in a forest.
c. *A bike plays a global map in a pot.

Two facts about language comprehension

i.) we have a potentially endless capacity to build the semantic representation of novel meaningful sentences

ii.) *ceteris paribus*, novel sentences (i.e., representing unexpected events) have a different cognitive status (i.e., they are processed differently) from familiar sentences

Language Comprehension (II)

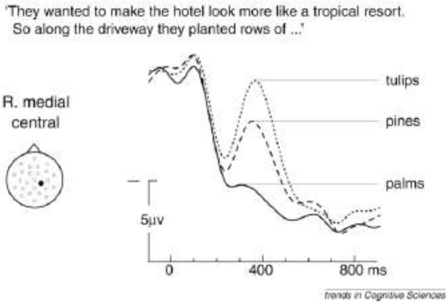
(1)
a. A student reads a book in a library.
b. A surfer reads a papyrus in a forest.
c. *A bike plays a global map in a pot.

Two facts about language comprehension

i.) we have a potentially endless capacity to build the semantic representation of novel meaningful sentences

ii.) *ceteris paribus*, novel sentences (i.e., representing unexpected events) have a different cognitive status (i.e., they are processed differently) from familiar sentences

Language Comprehension in the Brain



Sentences including possible but unexpected (novel) combinations of lexical items, evoke stronger N400 components in the ERP waveform than sentences with expected (non-novel) combinations (Kutas and Federmeier 2000, Baggio and Hagoort 2012)

A Balance between Storage and Computation

Baggio et al. (2012), "The processing consequences of compositionality", in M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford University Press

- Productivity entails that not everything can be stored in semantic memory, and that the brain is able to build semantic representations compositionally
- ERP data suggest that there is a large amount of stored knowledge in semantic memory about event contingencies and concept combinations (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more "distant" from the stored ones (e.g., novel combinations) require more cognitive effort to be interpreted

A Balance between Storage and Computation

Baggio et al. (2012), "The processing consequences of compositionality", in M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford University Press

- Productivity entails that not everything can be stored in semantic memory, and that the brain is able to build semantic representations compositionally
- ERP data suggest that there is a large amount of stored knowledge in semantic memory about event contingencies and concept combinations (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more "distant" from the stored ones (e.g., novel combinations) require more cognitive effort to be interpreted

A Balance between Storage and Computation

Baggio et al. (2012), "The processing consequences of compositionality", in M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford University Press

- Productivity entails that not everything can be stored in semantic memory, and that the brain is able to build semantic representations compositionally
- ERP data suggest that there is a large amount of stored knowledge in semantic memory about event contingencies and concept combinations (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more "distant" from the stored ones (e.g., novel combinations) require more cognitive effort to be interpreted

A Balance between Storage and Computation

Baggio et al. (2012), "The processing consequences of compositionality", in M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford University Press

- Productivity entails that not everything can be stored in semantic memory, and that the brain is able to build semantic representations compositionally
- ERP data suggest that there is a large amount of stored knowledge in semantic memory about event contingencies and concept combinations (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more "distant" from the stored ones (e.g., novel combinations) require more cognitive effort to be interpreted

Semantic Memory and Sentence Comprehension

Paczynski & Kuperberg (2012), “Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions”, *Journal of memory and Language*, 67: 426–448

- Comprehenders use different types of stored semantic information, including:
  - knowledge about the semantic relatedness between groups of concepts (e.g, *music*, *bass* and *guitarist* are semantically related to each other by sharing a common general schema)
  - structured knowledge about events, semantic roles and typical participants (e.g., knowing that a *bass* is more likely to be strummed by a *guitarist* than by a *drummer*)
- All these types of knowledge interact to predict the plausibility (expectancy) of an incoming word, given a preceding context (as reflected in N400 effects)

Semantic Memory and Sentence Comprehension

Paczynski & Kuperberg (2012), "Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions", *Journal of memory and Language*, 67: 426–448

- Comprehenders use different types of stored semantic information, including:
  - knowledge about the semantic relatedness between groups of concepts (e.g, *music*, *bass* and *guitarist* are semantically related to each other by sharing a common general schema)
  - structured knowledge about events, semantic roles and typical participants (e.g., knowing that a *bass* is more likely to be strummed by a *guitarist* than by a *drummer*)
- All these types of knowledge interact to predict the plausibility (expectancy) of an incoming word, given a preceding context (as reflected in N400 effects)

Generalized Event Knowledge (GEK)

McRae and Matsuki (2009), "People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible", *Language and Linguistics Compass*, 3:1417-1429

- Long-term semantic memory stores generalized knowledge about events and their participants (GEK)
- GEK derives from first-hand experience and from linguistic experience (e.g., from linguistic descriptions of events)
- Linguistic expressions are cues to activate various aspects of GEK stored in long-term memory

"Instrument nouns can cue certain types of eating, as in *eating with a fork* versus *eating with a stick*. Finally, event nouns like *breakfast* or location nouns like *cafeteria* cue specific types of eating scenarios" (McRae and Matsuki 2009: 1419)

"words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not *have* meaning; they are rather *cues* to meaning" (Elman 2014: 129)

Generalized Event Knowledge (GEK)

McRae and Matsuki (2009), "People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible", *Language and Linguistics Compass*, 3:1417-1429

- Long-term semantic memory stores generalized knowledge about events and their participants (GEK)
- GEK derives from first-hand experience and from linguistic experience (e.g., from linguistic descriptions of events)
- Linguistic expressions are cues to activate various aspects of GEK stored in long-term memory

"Instrument nouns can cue certain types of eating, as in *eating with a fork* versus *eating with a stick*. Finally, event nouns like *breakfast* or location nouns like *cafeteria* cue specific types of eating scenarios" (McRae and Matsuki 2009: 1419)

"words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not *have* meaning; they are rather *cues* to meaning" (Elman 2014: 129)

Generalized Event Knowledge (GEK)

McRae and Matsuki (2009), "People Use their Knowledge of Common Events to Understand Language, Do So as Quickly as Possible", *Language and Linguistics Compass*, 3:1417-1429

- Long-term semantic memory stores generalized knowledge about events and their participants (GEK)
- GEK derives from first-hand experience and from linguistic experience (e.g., from linguistic descriptions of events)
- Linguistic expressions are cues to activate various aspects of GEK stored in long-term memory

  "Instrument nouns can cue certain types of eating, as in *eating with a fork* versus *eating with a stick*. Finally, event nouns like *breakfast* or location nouns like *cafeteria* cue specific types of eating scenarios" (McRae and Matsuki 2009: 1419)

  "words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not *have* meaning; they are rather *cues* to meaning" (Elman 2014: 129)

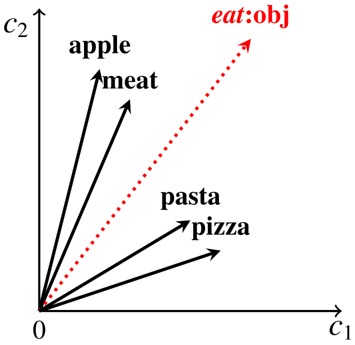
GEK and Thematic Fit

- Verb argument expectations are exploited by subjects during on line sentence processing to determine the plausibility of a noun as an argument of a verb (thematic fit)
  - cf. McRae *et al.* (1998), Kamide *et al.* (2003), among others

arrest a thief	highly prototypical
arrest a policeman	possible, but *less prototypical*
arrest a tree	impossible

Distributional Models of Thematic Fit

- Thematic fit judgments have been successfully modeled in distributional semantics
  - cf. Erk (2007), Baroni & Lenci (2010), Erk *et al.* (2010), Sayeed & Demberg, 2014; Sayeed et al., 2015; Greenberg et al., 2015; Sayeed et al., 2016; Tilk et al. 2016, Santus et al., 2017)



A Distributional Model of Sentence Comprehension

Chersoni, Lenci, Blache (2017), “Logical Metonymy in a Distributional Model of Sentence Comprehension”, *Proceedings *SEM 2017*: 168-177

- The distributional model is formed by a memory component and a unification component
  - cf. the Memory, Unification and Control (MUC) model by Hagoort (2016)
- The memory component stores Generalized Event Knowledge (*GEK*) modeled with distributional information extracted from large parsed corpora
- During sentence processing, lexical items (and constructions in general) activate portions of GEK, which are then unified to form a coherent semantic representation of the event expressed by the sentence
- Each semantic representation is associated with a compositional cost determining the sentence semantic complexity and depending on two factors:
  1. the availability and salience of “ready-to-use” event information already stored in GEK and cued by lexical items (constructions)
  2. the cost of unifying activated GEK into a coherent semantic representation, with the latter depending on the mutual semantic congruence of the events participants

A Distributional Model of Sentence Comprehension

Chersoni, Lenci, Blache (2017), "Logical Metonymy in a Distributional Model of Sentence Comprehension", *Proceedings *SEM 2017*: 168-177

- The distributional model is formed by a memory component and a unification component
  - cf. the Memory, Unification and Control (MUC) model by Hagoort (2016)
- The memory component stores Generalized Event Knowledge (*GEK*) modeled with distributional information extracted from large parsed corpora
- During sentence processing, lexical items (and constructions in general) activate portions of GEK, which are then unified to form a coherent semantic representation of the event expressed by the sentence
- Each semantic representation is associated with a compositional cost determining the sentence semantic complexity and depending on two factors:
  1. the availability and salience of "ready-to-use" event information already stored in GEK and cued by lexical items (constructions)
  2. the cost of unifying activated GEK into a coherent semantic representation, with the latter depending on the mutual semantic congruence of the events participants

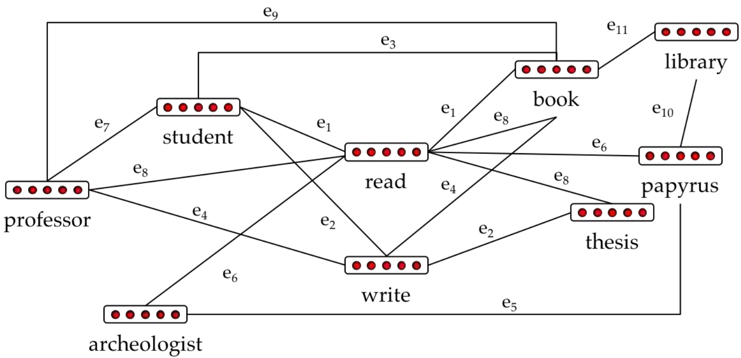
A Distributional Model of Sentence Comprehension

Chersoni, Lenci, Blache (2017), "Logical Metonymy in a Distributional Model of Sentence Comprehension", *Proceedings *SEM 2017*: 168-177

- The distributional model is formed by a memory component and a unification component
  - cf. the Memory, Unification and Control (MUC) model by Hagoort (2016)
- The memory component stores Generalized Event Knowledge (*GEK*) modeled with distributional information extracted from large parsed corpora
- During sentence processing, lexical items (and constructions in general) activate portions of GEK, which are then unified to form a coherent semantic representation of the event expressed by the sentence
- Each semantic representation is associated with a compositional cost determining the sentence semantic complexity and depending on two factors:
  1. the availability and salience of "ready-to-use" event information already stored in GEK and cued by lexical items (constructions)
  2. the cost of unifying activated GEK into a coherent semantic representation, with the latter depending on the mutual semantic congruence of the events participants

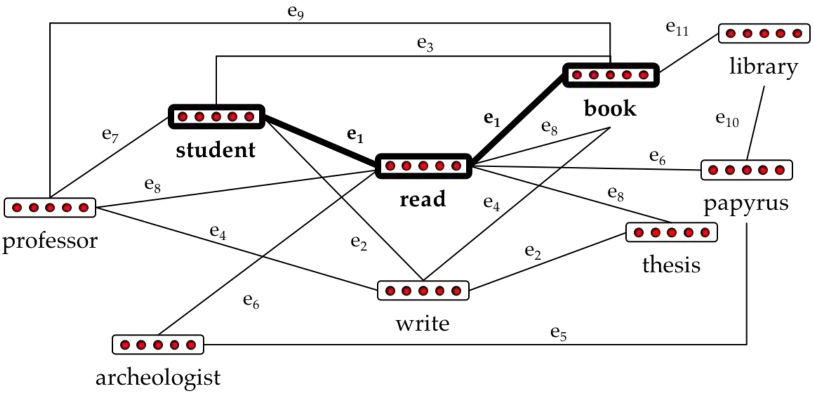
GEK and Distributional Semantics

- GEK is a deeply interrerelated network of events and participants, automatically extracted from the linguistic input (e.g. a parsed corpus)
  - nodes are distributional vectors of lexemes
  - edges correspond to relations between lexemes (e.g., thematic roles, syntagmatic associations, etc.) weighted with their statistical salience



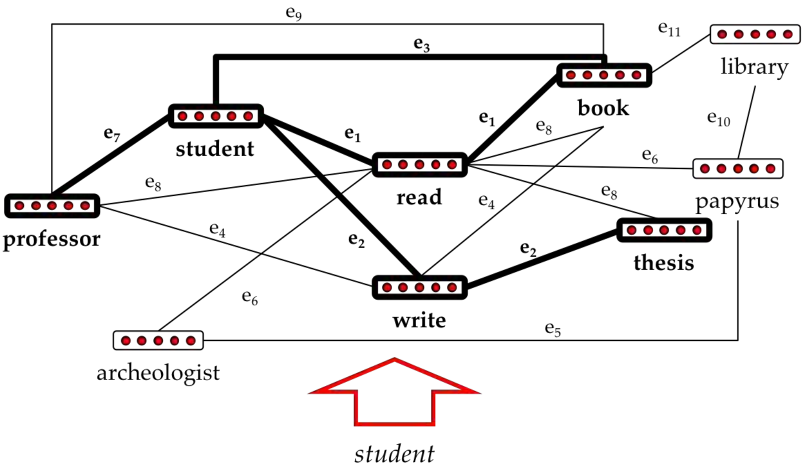
GEK and Distributional Semantics

- An event is a path in the GEK graph



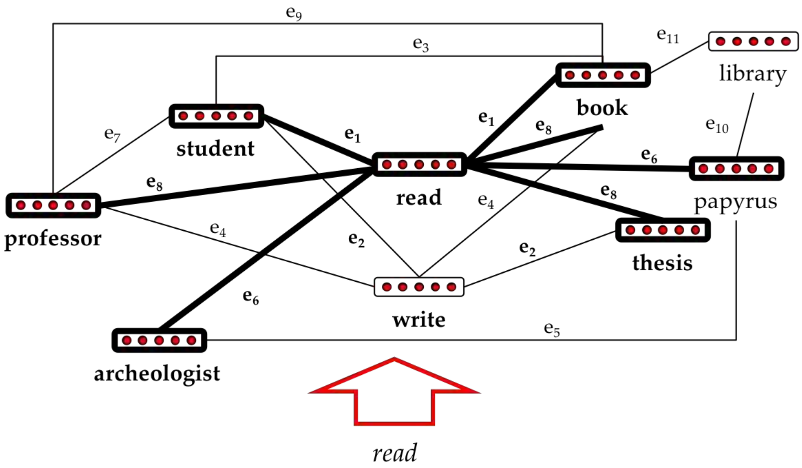
GEK and Distributional Semantics

- Lexical items and constructions cue (i.e., activate) portions of the GEK graph



GEK and Distributional Semantics

- Lexical items and constructions cue (i.e., activate) portions of the GEK graph



Distributional Semantic Composition

- Semantic composition is modeled as an event retrieval and construction process F, whose aim is to build a semantically coherent representation (SR) of a sentence by integrating the *GEK* cued by its elements
- Given an input sentence s, its interpretation INT(s) is the event that best explains its linguistic cues (Kuperberg 2016)
- INT(s) can be an event already stored in GEK and simply retrieved from it

  A student reads a book.
- ...or a new event constructed by linking together portions retrieved from GEK

  A surfer reads a papyrus.

Distributional Semantic Composition

- Semantic composition is modeled as an event retrieval and construction process F, whose aim is to build a semantically coherent representation (SR) of a sentence by integrating the *GEK* cued by its elements
- Given an input sentence s, its interpretation INT(s) is the event that best explains its linguistic cues (Kuperberg 2016)
- INT(s) can be an event already stored in GEK and simply retrieved from it

  A student reads a book.
- …or a new event constructed by linking together portions retrieved from GEK

  A surfer reads a papyrus.

Distributional Semantic Composition

- Semantic composition is modeled as an event retrieval and construction process F, whose aim is to build a semantically coherent representation (SR) of a sentence by integrating the *GEK* cued by its elements
- Given an input sentence s, its interpretation INT(s) is the event that best explains its linguistic cues (Kuperberg 2016)
- INT(s) can be an event already stored in GEK and simply retrieved from it

  A student reads a book.
- . . . or a new event constructed by linking together portions retrieved from GEK

  A surfer reads a papyrus.

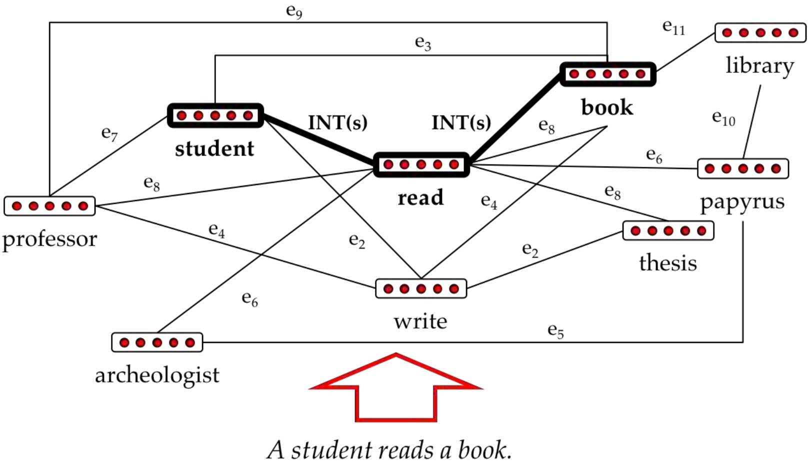
Distributional Semantic Composition

- Semantic composition is modeled as an event retrieval and construction process F, whose aim is to build a semantically coherent representation (SR) of a sentence by integrating the *GEK* cued by its elements
- Given an input sentence s, its interpretation INT(s) is the event that best explains its linguistic cues (Kuperberg 2016)
- INT(s) can be an event already stored in GEK and simply retrieved from it

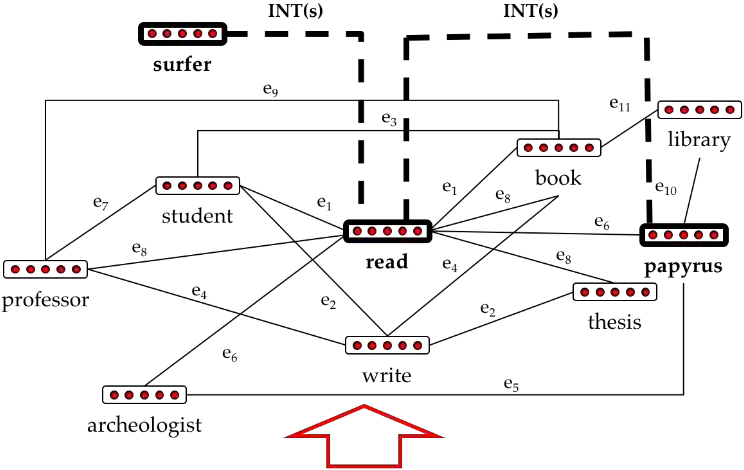
  A student reads a book.
- ...or a new event constructed by linking together portions retrieved from GEK

  A surfer reads a papyrus.

Distributional Semantic Composition



Distributional Semantic Composition



A surfer reads a papyrus.

Weighting Events during Semantic Interpretation

- While processing a sentence, the composition function weights events with respect to two dimensions:
  - the degree of activation by linguistic expressions (σ) to estimate the importance of "ready-to-use" event structures stored in GEK and retrieved during sentence processing
  - the internal semantic coherence (θ) of new events not stored in the memory component, and created with unification
- The joint effect of σ and θ captures the "balance between storage and computation" driving sentence processing

Weighting Events during Semantic Interpretation

- While processing a sentence, the composition function weights events with respect to two dimensions:
  - the degree of activation by linguistic expressions (σ) to estimate the importance of "ready-to-use" event structures stored in GEK and retrieved during sentence processing
  - the internal semantic coherence (θ) of new events not stored in the memory component, and created with unification
- The joint effect of σ and θ captures the "balance between storage and computation" driving sentence processing

Weighting Events during Semantic Interpretation

- While processing a sentence, the composition function weights events with respect to two dimensions:
  - the degree of activation by linguistic expressions (σ) to estimate the importance of “ready-to-use” event structures stored in GEK and retrieved during sentence processing
  - the internal semantic coherence (θ) of new events not stored in the memory component, and created with unification
- The joint effect of σ and θ captures the “balance between storage and computation” driving sentence processing

Weighting Events during Semantic Interpretation

- The score σ is a linear function of the event weights cued by linguistic items
  - events that are cued by more linguistic constructions in a sentence should incrementally increase their salience
- The score θ assumes that the internal coherence of an event depends on the mutual typicality among the components of an event
  - e.g. a surfer is similar enough to typical readers (e.g., s/he is animate), but s/he is not similar to typical "papyrus-readers"
- Semantic typicality is measured with thematic fit cosine (cf. Erk et al. 2010, Baroni and Lenci 2010, Chersoni et al. 2017), using the distributional representations of the GEK nodes.

Weighting Events during Semantic Interpretation

- The score σ is a linear function of the event weights cued by linguistic items
  - events that are cued by more linguistic constructions in a sentence should incrementally increase their salience
- The score θ assumes that the internal coherence of an event depends on the mutual typicality among the components of an event
  - e.g. a surfer is similar enough to typical readers (e.g., s/he is animate), but s/he is not similar to typical "papyrus-readers"
- Semantic typicality is measured with thematic fit cosine (cf. Erk et al. 2010, Baroni and Lenci 2010, Chersoni et al. 2017), using the distributional representations of the GEK nodes.

Weighting Events during Semantic Interpretation

- The score σ is a linear function of the event weights cued by linguistic items
  - events that are cued by more linguistic constructions in a sentence should incrementally increase their salience
- The score θ assumes that the internal coherence of an event depends on the mutual typicality among the components of an event
  - e.g. a surfer is similar enough to typical readers (e.g., s/he is animate), but s/he is not similar to typical "papyrus-readers"
- Semantic typicality is measured with thematic fit cosine (cf. Erk et al. 2010, Baroni and Lenci 2010, Chersoni et al. 2017), using the distributional representations of the GEK nodes.

Distributional Semantic Composition

Given an input sentence s, its interpretation INT(s) is the event e_k with the highest semantic composition weight (SCW):

$$\mathrm{INT}(s) = \operatorname*{argmax}_e(\mathrm{SCW}(e)) \tag{1}$$

$$\mathrm{SCW}(e) = \theta_e + \sigma_e \tag{2}$$

Semantic Composition Cost

- The semantic composition cost of a sentence s is inversely related to the SCW of the event representing its interpretation:

$$\mathrm{SemComp}(s) = \frac{1}{\mathrm{SCW}(\mathrm{INT}(s))} \tag{3}$$

  - the less internally coherent is the event represented by the sentence and the less strong is its activation by the lexical items, the more the unification is cognitively expensive and the sentence semantically complex

Modelling Cognitive Data

- Context-sensitive argument typicality
  - Bicknell K. *et al.* (2010), "Effects of event knowledge in processing verbal arguments", *Journal of Memory and Language*, 63: 489-505
- Logical metonymy (coercion, enriched composition)
  - McElree B. *et al.* (2001), "Reading time evidence for enriching composition", *Cognition*, 78: B17–B25
  - Traxler, M. *et al.* (2002), "Coercion in sentence processing: evidence from eye-movements and self-paced reading", *Journal of Memory and Language*, 47: 530–547
- Selectional preference violation
  - Warren, T. *et al.* (2015), "Comprehending the impossible: what role do selectional restriction violations play?", *Language, Cognition and Neuroscience*, 30: 932–939

Chersoni E., Lenci A., Blache P. (2017), "Logical Metonymy in a Distributional Model of Sentence Comprehension", *Proceedings *SEM 2017)*.

Chersoni E., Santus E., Blache P., Lenci A. (2017),"Is Structure Necessary for Modeling Argument Expectations in Distributional Semantics?", *Proceedings of IWCS 2017*.

Santus E., Chersoni E., Lenci A., Blache P. (2017), "Measuring Thematic Fit with Distributional Feature Overlap", *Proceedings of EMNLP 2017*.

Extracting GEK

- Events were extracted from the British National Corpus (BNC), the Reuters Corpus vol.1 (RCV1), the ukWaC and the Wackypedia Corpus
    - events are formed by the verb and its direct dependencies: subject (NSUBJ), direct object (DOBJ), indirect object (IOBJ) and a generic prepositional complement relation (PREPCOMP)
    - 4,204,940 extracted events (including schematic ones)
- Each verb and noun occurring in these event structures was represented with a distributional vector in a syntax-based DSM using as contexts the extracted dependencies

Context-sensitive argument typicality

Bicknell K. *et al.* (2010), "Effects of event knowledge in processing verbal arguments", *Journal of Memory and Language*, 63: 489-505

- The Bicknell dataset includes 100 pairs of sentences (superset of the dataset used in Bicknell et al. 2010)
- Each pair contains a congruent and an incongruent sentence, that differ for the object typicality, but not for the subject one

(1) The **journalist** checked the **spelling** of his latest report (congruent)

(2) The **mechanic** checked the **spelling** of his latest report (incongruent)

The Bicknell Dataset

Bicknell K. *et al.* (2010), "Effects of event knowledge in processing verbal arguments", *Journal of Memory and Language*, 63: 489-505

- Self-paced reading and ERP studies show that the the typicality of a verb direct object depends on the subject argument

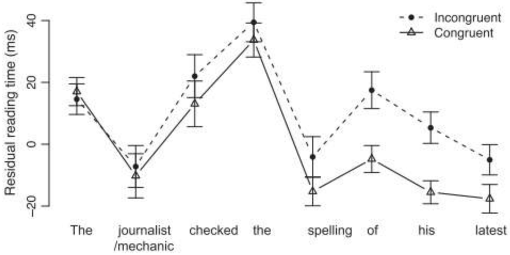
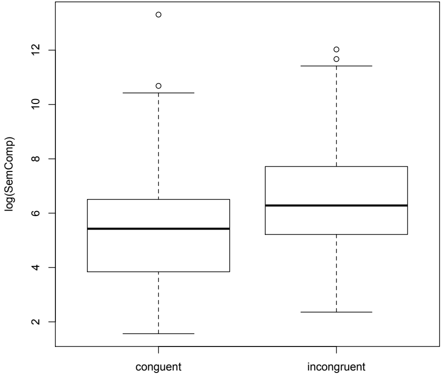


Fig. 1. Mean residual reading times. Error bars show one standard error above and below the mean, calculated by participants.

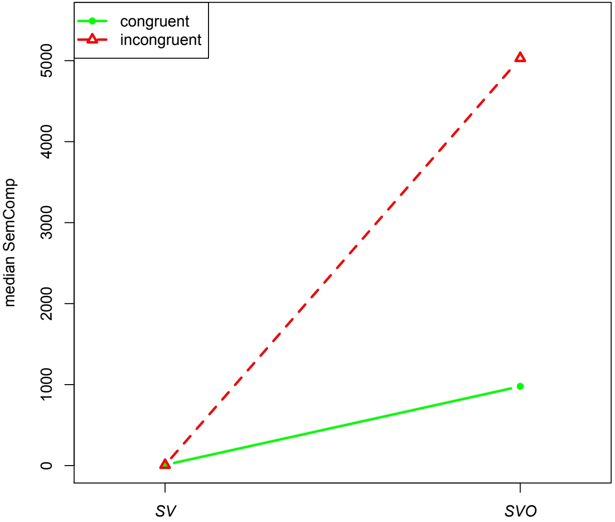
Results

- The semantic composition costs assigned to congruent sentences by the model are significantly lower than the scores assigned to incongruent sentences



Wilcoxon rank sum test: W = 2732, p-value < 0.001 (model coverage: 91 sentences)

The Dynamics of Compositional Cost



Distributional Representations

- Language shapes our semantic representations, although the contribution of linguistic experience vis-à-vis other kinds of non-linguistic inputs is an empirical question that is widely debated in cognitive science (Dove 2014), Louwerse (2011), Vigliocco et al.(2009)
- Under various respects, distributional semantics still offer a coarse-grained view of meaning, and many aspects are left out of the picture
  - e.g., the lexicon is often regarded as the “bottleneck” for symbolic models, but compositionality is surely the “bottleneck” for distributional semantics
- The continuous and distributed nature of distributional representations offers the opportunity to
  - tackle the variability, gradeness, and context-dependence of lexical meaning
  - learn such representations from data
  - provide new bridges with neurocognitive models of semantic memory and language processing

Distributional Representations

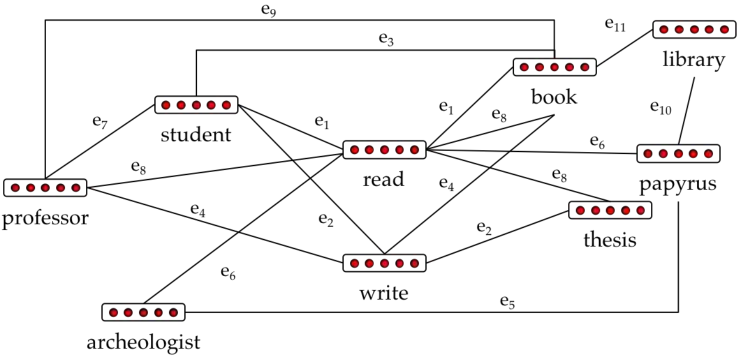
- Language shapes our semantic representations, although the contribution of linguistic experience vis-à-vis other kinds of non-linguistic inputs is an empirical question that is widely debated in cognitive science (Dove 2014), Louwerse (2011), Vigliocco et al.(2009)
- Under various respects, distributional semantics still offer a coarse-grained view of meaning, and many aspects are left out of the picture
  - e.g., the lexicon is often regarded as the "bottleneck" for symbolic models, but compositionality is surely the "bottleneck" for distributional semantics
- The continuous and distributed nature of distributional representations offers the opportunity to
  - tackle the variability, gradeness, and context-dependence of lexical meaning
  - learn such representations from data
  - provide new bridges with neurocognitive models of semantic memory and language processing

Distributional Representations

- Language shapes our semantic representations, although the contribution of linguistic experience vis-à-vis other kinds of non-linguistic inputs is an empirical question that is widely debated in cognitive science (Dove 2014), Louwerse (2011), Vigliocco et al.(2009)
- Under various respects, distributional semantics still offer a coarse-grained view of meaning, and many aspects are left out of the picture
  - e.g., the lexicon is often regarded as the “bottleneck” for symbolic models, but compositionality is surely the “bottleneck” for distributional semantics
- The continuous and distributed nature of distributional representations offers the opportunity to
  - tackle the variability, gradeness, and context-dependence of lexical meaning
  - learn such representations from data
  - provide new bridges with neurocognitive models of semantic memory and language processing

Distributional Semantics and Sentence Processing

- Distributional information is not only relevant to build vector representations for single lexemes, but also to model the network of associations among lexemes forming the events and situations (GEK) stored in semantic memory and crucial for language processing



Associative Networks of (Distributional) Vectors

Binder, J. R. (2016). "In defense of abstract conceptual representations". *Psychonomic Bulletin & Review* 23, 1096–1108

"Abstract representations in the brain arise from a process of hierarchical conjunctive coding, and it is their combinatorial nature that is important rather than their abstractness per se.

A related and equally ubiquitous phenomenon for which [crossmodal conjunctive representations] CCRs provide a much-needed explanation is thematic association. Consider the statement "The boy walked his dog in the park." The inference that the dog is likely wearing a leash cannot be made purely on the basis of the sensory-motor features of *dog*, *walk*, *park*, or *leash*. Rather, the leash is a thematic or situation-specific association based on co-occurrence experiences.

CCRs solve this problem by providing highly abstract conceptual representations activated by conjunctions of features, which can then "wire together" with other highly abstract conceptual representations with which they co-occur."

Distributional Semantics and Sentence Processing

- Distributional information stored in GEK is retrieved and combined during language comprehension, allowing humans to:
  - make predictions and generating expectancies about incoming events and participants
  - draw inferences (e.g., filling missing details about the described event)
- Understanding a sentence involves retrieving stored events and constructing new events
  - this process may cross the classical divide between stored – idiomatic vs. constructed – compositional (cf. Jackendoff 2002, 2013, among many others)

Distributional Semantics and Sentence Processing

- Distributional information stored in GEK is retrieved and combined during language comprehension, allowing humans to:
  - make predictions and generating expectancies about incoming events and participants
  - draw inferences (e.g., filling missing details about the described event)
- Understanding a sentence involves retrieving stored events and constructing new events
  - this process may cross the classical divide between stored – idiomatic vs. constructed – compositional (cf. Jackendoff 2002, 2013, among many others)

Distributional Semantics and Sentence Processing

- The interpretation of novel sentences (i.e., productivity) is obtained by retrieving and combining stored information to build the representation of new events
- The salience of a new event is a function of its internal semantic coherence, which in turns depends on its similarity to stored events
- Distributional representations allow us to measure the similarity of new items to those already stored in semantic memory
- Language productivity can be conceived as the capacity to adapt our knowledge stored in semantic memory to novel situations

Productivity is adaptation and adaptation is by similarity

Distributional Semantics and Sentence Processing

- The interpretation of novel sentences (i.e., productivity) is obtained by retrieving and combining stored information to build the representation of new events
- The salience of a new event is a function of its internal semantic coherence, which in turns depends on its similarity to stored events
- Distributional representations allow us to measure the similarity of new items to those already stored in semantic memory
- Language productivity can be conceived as the capacity to adapt our knowledge stored in semantic memory to novel situations

Productivity is adaptation and adaptation is by similarity

Distributional Semantics and Sentence Processing

- The interpretation of novel sentences (i.e., productivity) is obtained by retrieving and combining stored information to build the representation of new events
- The salience of a new event is a function of its internal semantic coherence, which in turns depends on its similarity to stored events
- Distributional representations allow us to measure the similarity of new items to those already stored in semantic memory
- Language productivity can be conceived as the capacity to adapt our knowledge stored in semantic memory to novel situations

Productivity is adaptation and adaptation is by similarity

Distributional Semantics and Sentence Processing

- The interpretation of novel sentences (i.e., productivity) is obtained by retrieving and combining stored information to build the representation of new events
- The salience of a new event is a function of its internal semantic coherence, which in turns depends on its similarity to stored events
- Distributional representations allow us to measure the similarity of new items to those already stored in semantic memory
- Language productivity can be conceived as the capacity to adapt our knowledge stored in semantic memory to novel situations

Productivity is adaptation and adaptation is by similarity

The Project Team

CoLing Lab
Computational Linguistics Laboratory

Patrick Jeuniaux

Ludovica Pannitto

Paolo Vassallo

BLRI
Brain & Language Research Institute

Philippe Blache

Emmanuele Chersoni

Merci!!!

Grazie!!!

Thank you!!!