Event Knowledge in Compositional Distributional Semantics

Ludovica Pannitto
Master Thesis in Digital Humanities - Language Technologies, University of Pisa
Supervisor: Alessandro Lenci

November 20, 2019

Contents

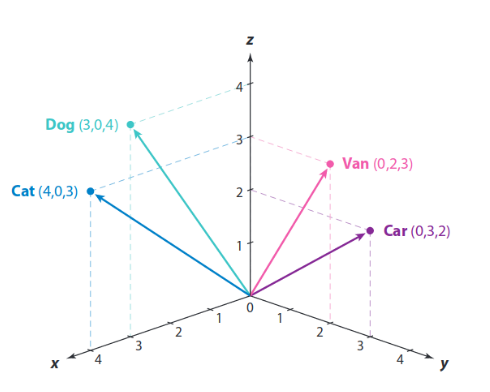
Aim: investigate the use of **distributional methods** in a model of **compositional meaning** that is both **linguistically motivated** and **cognitively inspired**.

- Background
    - Compositional Distributional Semantics
    - Generalized Event Knowledge
    - MUC: neurobiological model of language and processing
- Model
    - Description
    - Evaluation
- Discussion
    - Error analysis

Background

Compositional Distributional Semantics (1): DH

"what people know when they know a word is not how to recite its dictionary definition – they know how to use it (when to produce it and how to understand it) in everyday discourse" (Miller e Charles 1991)



Compositional Distributional Semantics (2)

Composing word representations into larger phrases and sentences notoriously represents a big challenge for distributional semantics[1].

Various approaches have been proposed ranging from simple arithmetic operations on word vectors, to algebraic compositional functions on higher-order objects, as well as neural networks approaches[2].

Vector addition still shows reasonable performances overall[3], its success being quite puzzling from the linguistic and cognitive point of view.

---

[1] Lenci 2018

[2] Mitchell e Lapata 2008; Coecke, Clark e Sadrzadeh 2010; Socher, Manning e Ng 2010; Mikolov et al. 2013; Baroni, Bernardi e Zamparelli 2014

[3] or at least it was when we started this work

Generalized Event Knowledge (1): Accettability vs. Plausibility

The problem of compositionality has for long been addressed as a distinction between *possible* and *impossible* sentences:

(1) The musician plays the flute in the theater.

(2) * The nominative plays the global map in the pot.

The first class subsumes a great amount of phenomena, coalescing **typical** and **atypical** sentences:

(3) The gardener plays the castanets in the cave.

Generalized Event Knowledge (2)

Psycholinguistic evidence shows that lexical items activate a great amount of **generalized event knowledge** (GEK)[4] about **typical** events, and that this knowledge is crucially exploited **during** online language processing, constraining the speakers' expectations about upcoming linguistic input[5].

(4) The **man** arrested...*by the police*

(5) The **cop** arrested...*a man yesterday*

[4] Elman 2011; Hagoort e Berkum 2007; Hare et al. 2009
[5] McRae e Matsuki 2009

Generalized Event Knowledge (3): the lexicon

The mental lexicon is organized as a **network of mutual expectations** which are in turn able to influence comprehension.

Sentence comprehension is phrased as the **identification of the event that best explains the linguistic cues used in the input**[6] .

[6] Kuperberg e Jaeger 2016

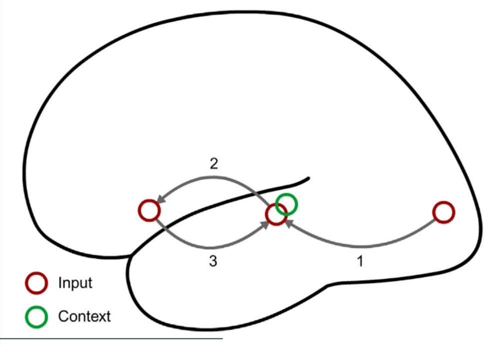
Memory, Unification and Control

The architecture is based on the Memory, Unification and Control (MUC) model[7]:

Memory - linguistic knowledge stored in long-term memory

Unification - constraint-based assembly of linguistic items in working memory

Control - relating language to joint action and interaction



[7] Hagoort 2015

Model

The purpose is to **integrate** vector addition with **Generalized Event Knowledge** activated by lexical items.

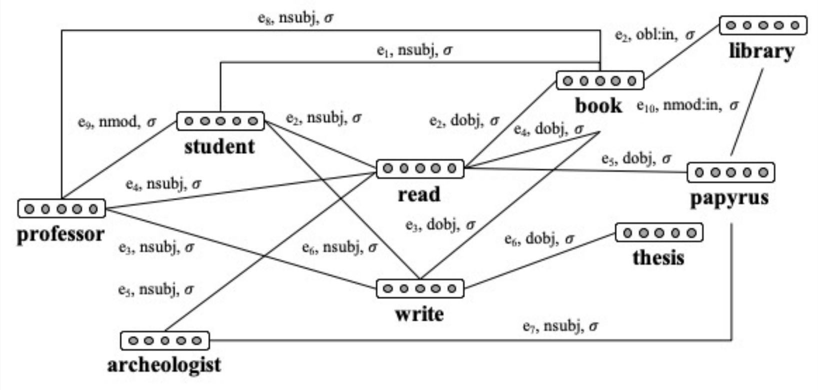
It is directly inspired by previous models[8] and consists of two components:

Distributional Event Graph (DEG) - embeddings in a network of syntagmatic relations, modeling a fragment of semantic memory activated by lexical units;

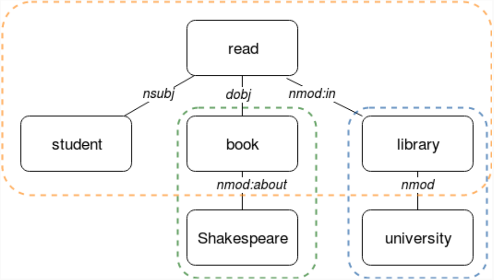
Meaning Composition Function - dynamically builds a **structured object** using information activated from DEG through lexical items.

[8] Chersoni, Lenci e Blache 2017

DEG (1) at a glance

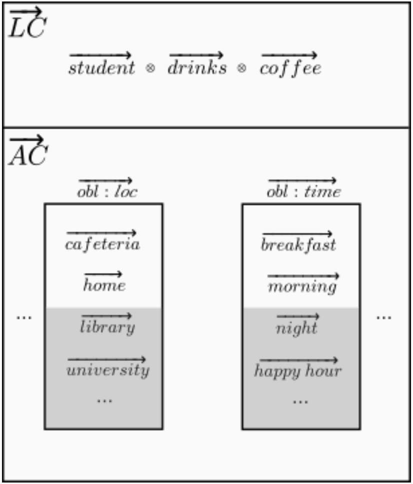


DEG (2): construction

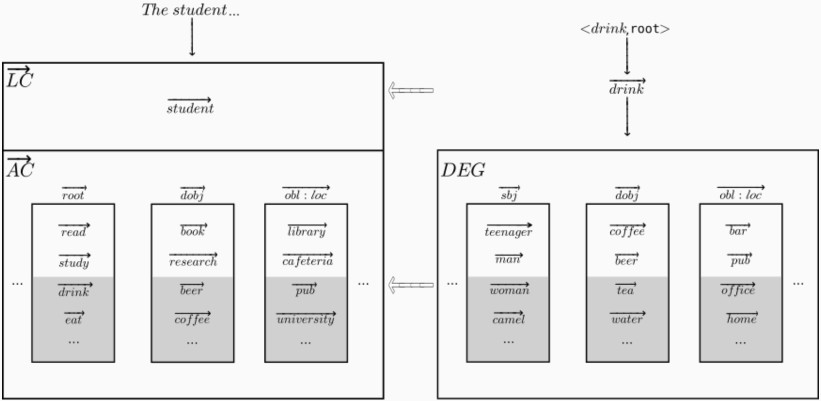


We assume a broad notion of event, corresponding to **any configuration of entities, actions, properties, and relationships**, also schematic or **underspecified**.
Events are cued by all the potential participants, depending on the statistical association between the event and the participant.

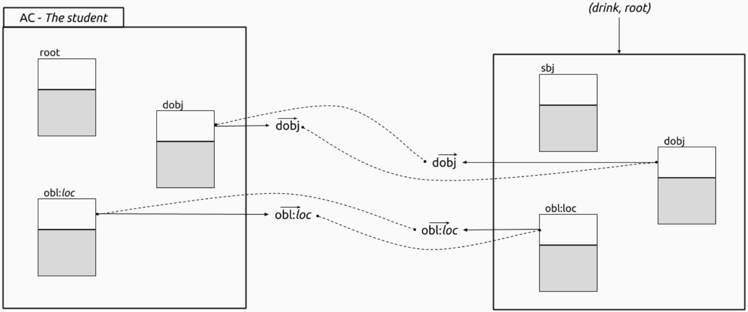
The student drinks coffee (1)



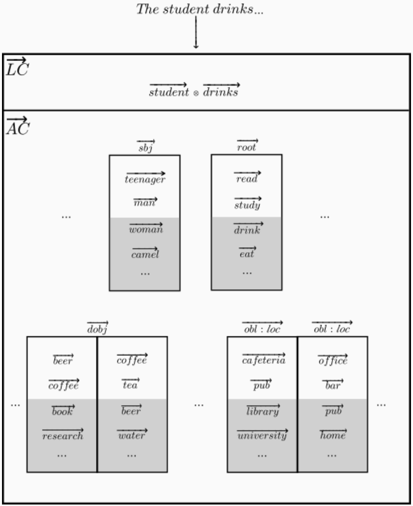
The student… (2)



The student... (3): weighting process



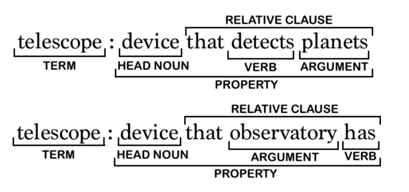
The student drinks ... (4)



Evaluation

Datasets

RELPRON[9]:



TSS dataset[10]:

(6) a. government use power
b. authority exercise influence

(7) a. team win match
b. design reduce amount

[9] 518 semi-automatically created pairs, Rimell et al. 2016.
[10] 108 pairs of sentences annotated with human judgments, Kartsaklis e Sadrzadeh 2014.

Each item was represented as a triplet:

RELPRON - $(hn, r), (w_1, nsubj/root), (w_2, root/dobj)$

TSS - $(w_1, nsubj), (w_2, root), (w_3, dobj)$

Each item was represented as a triplet:

RELPRON - $(hn, r), (w_1, nsubj/root), (w_2, root/dobj)$

TSS - $(w_1, nsubj), (w_2, root), (w_3, dobj)$

We tested 6 (7 for TSS) settings, containing all the possible combinations or arguments.

Each item was represented as a triplet:

RELPRON - $(hn, r), (w_1, nsubj/root), (w_2, root/dobj)$

TSS - $(w_1, nsubj), (w_2, root), (w_3, dobj)$

We tested 6 (7 for TSS) settings, containing all the possible combinations or arguments.

For each model, we built a semantic representation $\text{SR} = (\text{LC}, \text{AC})$, where:

LC is built through vector addition and represents our baseline

AC is limited to the overtly filled participants and is used as a representation of Generalized Event Knowledge

Task

RELPRON - for each target noun, we produced a ranking over all the available properties and computed Mean Average Precision

$$s = cos(\overrightarrow{target}, \overrightarrow{LC}) + cos(\overrightarrow{target}, \overrightarrow{AC}) \tag{1}$$

TSS - we evaluated the correlation of our scores with human ratings with Spearman's ρ

$$s = cos(\overrightarrow{LC_1}, \overrightarrow{LC_2}) + cos(\overrightarrow{AC_1}, \overrightarrow{AC_2}) \tag{2}$$

Results - RELPRON - MAP scores

	RELPRON		
	LC[11]	AC	LC+AC
verb	0,18	0,18	**0,20**
arg	0,34	0,34	**0,36**
hn+verb	0,27	0,28	**0,29**
hn+arg	0,47	0,45	**0,49**
verb+arg	**0,42**	0,28	0,39
hn+verb+arg	0,51	0,47	**0,55**

[11] VECTOR ADDITION ONLY

Results - TSS - ρ scores

	transitive sentences dataset		
	LC[12]	AC	LC+AC
sbj	0.432	0.475	0.**482**
root	0.525	0.547	0.**555**
obj	0.628	0.537	0.**637**
sbj+root	0.**656**	0.622	0.648
sbj+obj	0.653	0.605	0.**656**
root+obj	0.732	0.696	0.**750**
sbj+root+obj	0.732	0.686	0.**750**

[12] VECTOR ADDITION ONLY

Error Analysis

RELPRON plausibility

Target noun: navy

- organization that general commands
- organization that soldier serves
- organization that uses submarine
- organization that blockades port

RELPRON plausibility

Target noun: navy

- organization that general commands
- organization that soldier serves
- **organization that uses submarine**
- **organization that blockades port**

We collected human similarity judgements for highly typical paraphrases and atypical (random) paraphrases.

RELPRON plausibility

Target noun: **navy**

- organization that general commands
- organization that soldier serves
- organization that uses submarine
- organization that blockades port

We collected human similarity judgements for highly typical paraphrases and atypical (random) paraphrases.

	RELPRON items			RANDOM items		
	LC	AC	LC+AC	LC	AC	LC+AC
verb	0,06	**0,08**	0,07	0,26	0,23	**0,27**
arg	**0,22**	0,16	0,20	0,27	**0,32**	0,31
hn+verb	0,01	**0,04**	0,02	0,13	**0,21**	0,18
hn+arg	**0,18**	0,15	**0,18**	0,21	**0,28**	0,26
verb+arg	**0,20**	0,06	0,14	0,31	0,30	**0,33**
hn+verb+arg	**0,16**	0,09	0,14	0,25	0,24	**0,26**

* scores are expressed as ρ correlations

subject vs. *object* relative clauses

Subject realtive clauses perform generally worse than *object* clauses, especially in the *verb + arg* setting.

AC	verb	arg	hn+verb	hn+arg	verb+arg	hn+verb+arg
subject	0,19	0,41	0,29	0,47	0,22	0,48
object	0,19	0,34	0,29	0,51	0,38	0,52
Δ	0,00	0,06	0,00	-0,04	**-0,16**	-0,04

The model processes items in **linear order**: the *verb+arg* setting works differently when applied to subject clauses than to object clauses.
In the subject case the verb is found first, and then its expectations are used to re-rank the object ones. In the object case things proceed the opposite way.

Wrap-up

We provided a basic implementation of a meaning composition model, which aims at being **incremental** and **cognitively plausible**.

While still relying on vector addition, our results suggest that distributional vectors do not encode sufficient information about event knowledge, and that, in line with psycholinguistic results, activated GEK plays an important role in building semantic representations during online sentence processing.

Thank you! :)

- Emmanuele Chersoni, Enrico Santus, Ludovica Pannitto, Alessandro Lenci, Philippe Blache, Chu-Ren Huang (2019), “A structured distributional model of sentence meaning and processing“, Natural Language Engineering, 25: 483-502
- Ludovica Pannitto, Alessandro Lenci, *to appear*, ”Event Knowledge in Compositional Distributional Semantics“, Italian Journal of Computational Linguistics

References

Baroni, Marco, Raffaela Bernardi e Roberto Zamparelli (2014). «Frege in Space: A Program of Compositional Distributional Semantics». In: *Linguistic Issues in Language Technology* 9.6, pp. 5–110.

Chersoni, Emmanuele, Alessandro Lenci e Philippe Blache (2017). «Logical Metonymy in a Distributional Model of Sentence Comprehension». In: *Sixth Joint Conference on Lexical and Computational Semantics (* SEM 2017)*, pp. 168–177.

Coecke, Bob, Stephen Clark e Mehrnoosh Sadrzadeh (2010). *Mathematical Foundations for a Compositional Distributional Model of Meaning*. Rapp. tecn.

Elman, Jeffrey L (2011). «Lexical knowledge without a lexicon?». In: *The mental lexicon* 6.1, pp. 1–33.

Hagoort, Peter (2015). «MUC (memory, unification, control): A model on the neurobiology of language beyond single word processing». In: *Neurobiology of language*. Elsevier, pp. 339–347.

Hagoort, Peter e Jos van Berkum (2007). «Beyond the sentence given». In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1481, pp. 801–811.

Hare, Mary et al. (2009). «Activating event knowledge». In: *Cognition* 111.2, pp. 151–167.

Kartsaklis, Dimitri e Mehrnoosh Sadrzadeh (2014). «A Study of Entanglement in a Categorical Framework of Natural Language». In: *Proceedings of the 11th Workshop on Quantum Physics and Logic (QPL)*. Kyoto ,Japan.

Kuperberg, Gina R e T Florian Jaeger (2016). «What do we mean by prediction in language comprehension?». In: *Language, cognition and neuroscience* 31.1, pp. 32–59.

Lenci, Alessandro (2018). «Distributional Models of Word Meaning». In: *Annual Review of Linguistics* 4, pp. 151–171.

McRae, Ken e Kazunaga Matsuki (2009). «People use their knowledge of common events to understand language, and do so as quickly as possible». In: *Language and linguistics compass* 3.6, pp. 1417–1429.

Mikolov, Tomas et al. (2013). «Distributed representations of words and phrases and their compositionality». In: *Advances in neural information processing systems*, pp. 3111–3119.

Miller, George A e Walter G Charles (1991). «Contextual correlates of semantic similarity». In: *Language and cognitive processes* 6.1, pp. 1–28.

Mitchell, Jeff e Mirella Lapata (2008). «Vector-based Models of Semantic Composition.». In:

Rimell, Laura et al. (2016). «RELPRON: A relative clause evaluation data set for compositional distributional semantics». In: *Computational Linguistics* 42.4, pp. 661–701.

Socher, Richard, Christopher D Manning e Andrew Y Ng (2010). «Learning continuous phrase representations and syntactic parsing with recursive neural networks». In: *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*. Vol. 2010, pp. 1–9.