

Capturing horizontal links in a construction network using semantic vector spaces

Guillaume Desagulier^{1, 2, 3}

¹Université Paris 8

²MoDyCo (UMR 7114), CNRS, Université Paris Nanterre

³Institut Universitaire de France

Modelling constructional variation and change
University of Zurich - November 15th, 2021

plan

- 1 introduction
- 2 research context
- 3 the IL construction
- 4 discussion

the constructicon

The constructicon consists of nodes and links :

- **nodes** : form-meaning pairings that are characterized by various degrees of schematicity
- **links**
 - vertical links**, which unite the lower-level instantiations to the higher-level schemas in a bottom-up fashion
 - horizontal links**, which unite nodes at the same level of schematicity (DIESSEL 2019).

changing constructions

In DxCG, construction-network changes are recognized to fall into three categories :

- node creation ('constructionalization')
- node-internal changes ('constructional change')
- node-external changes (SMIRNOVA et SOMMERER 2020 ; E. C. TRAUGOTT et TROUSDALE 2013)

What about the changes that affect the links ?

- often, vertical links are analyzed from the angle of **productivity** and **schematicity** via distributional analysis

horizontal links

- horizontal links deserve more attention than they have received
- it is only by studying horizontal links that we can understand phenomena such as constructional competition or division of labor (ZEHENTNER et E. TRAUGOTT 2020)

goal

I want to show how horizontal links can be captured and described using distributional semantic modelling

Distributional semantic models

- computational implementations of the distributional hypothesis : words that occur in similar contexts tend to have similar meanings (FIRTH 1957; HARRIS 1954; MILLER et CHARLES 1991)
- used to produce semantic representations of words from co-occurrence matrices
- originally, a co-occurrence matrix is populated with **frequency counts** and each row is an array of such frequencies, also known as a **vector**
- they allow the linguist to hold the form constant and track meaning shifts at regular intervals in a diachronic corpus (HAMILTON et al. 2016; KULKARNI et al. 2015; SAGI et al. 2011).

HILPERT (2016)

- goal : justify a constructional reading of English modals
- how : builds a semantic vector space with the collocates of the most frequent verbs that occur with *may* in a 50-M word sample of the COCA
- diachronic frequency information from the COHA is then projected onto the reference semantic vector space in the form of contour plots at regular intervals (1800s–1860s, 1870s–1920s, 1930s–1990s)
- *may* entertains a complex network of associations with the lexical verbs that it governs and that it has shifted away from the expression deontic modal meanings towards epistemic meanings and a higher degree of informativeness

PEREK (2016)

- goal : show that the distributional semantic approach to semantic similarity can be applied successfully to syntactic productivity in diachrony
- how : assess the shifting structure of the semantic domain of the *hell*-construction (*V the hell out of NP*) at 4 points in time in the COHA (1930s-- 1940s, 1950s--1960s, 1970s--1980s, and 1990s--2000s)

PEREK (2018)

- goal : investigate change in the way-construction between the 1830s and the 2000s
- how : the reference vector space is built with data from the COHA
- the 3 senses of the way-construction (path-creation, manner, and incidental-action) have gained in semantic diversity
- the verb slot of the motion component has become more schematic and more productive

count-based DSMs



the vector space is then augmented with diachronic frequency information (words are plotted based on period-relevance, or density curves are plotted)

count-based DSMs

- HILPERT (2016) and PEREK (2016, 2018) use DSMs that rely on count models (whose vectors are generated from co-occurrence counts)
- the resulting vectors are **long and sparse**, which is likely to affect the quality of the vector representations, even after weighting and dimensionality reduction (BARONI et al. 2014)

predictive DSMs

Predictive models such as `word2vec` are inspired by neural language models (BENGIO et al. 2003; COLLOBERT et al. 2011).

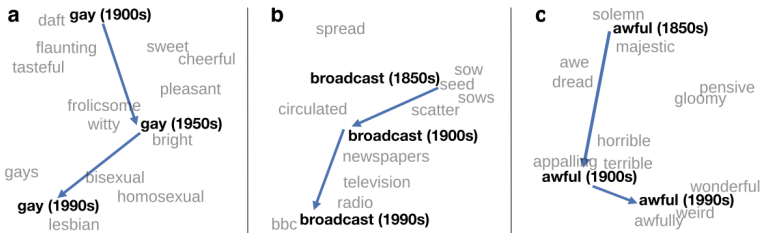
Instead of counting how often a collocate \mathbf{c} occurs near a target word \mathbf{w} , predictive models estimate the probability of finding \mathbf{c} near \mathbf{w} .

The resulting vectors are :

- relatively **short** (from 50 to 1000 dimensions)
- **dense** (no cells with zeros).

HAMILTON et al. (2016)

- Skip-Gram with Negative Sampling (**SGNS**), a model of the **word2vec** toolkit (Mikolov et al. 2013), outperforms PPMI and SVD in the discovery of new shifts and the visualization of changes.
- **Histwords** is a collection of tools and datasets for Python. Its goal is to quantify semantic change by evaluating word embeddings
- works well with SGNS but when applied to conveniently selected lexemes

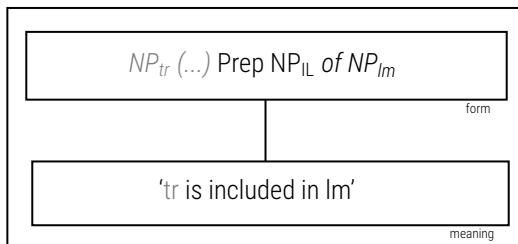


The IL construction

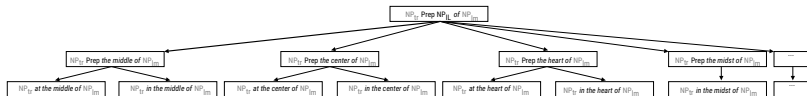
- (1) He stops suddenly **in the middle of** the stage and seems to consider.
(1815-FIC-FalseShame)
- (2) Marvin walked to the chalk mark **in the center of** the ring.
(1934-FIC-Captain Caution)
- (3) (...) we were **in the heart of** Norwalk. (1827-FIC-Novels)
- (4) We see St Eustace praying **in the midst of** the river.
(1980-FIC-RiddleyWalker)

These patterns denote a **relationship of inclusion** between a trajector and a landmark.

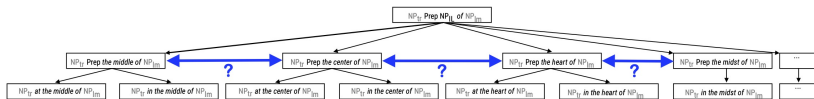
The IL construction



The IL construction



The IL construction



Data

- 15,782 tokens of the IL construction were retrieved from the COHA

$$N_{middle} = 5821$$

$$N_{midst} = 4243$$

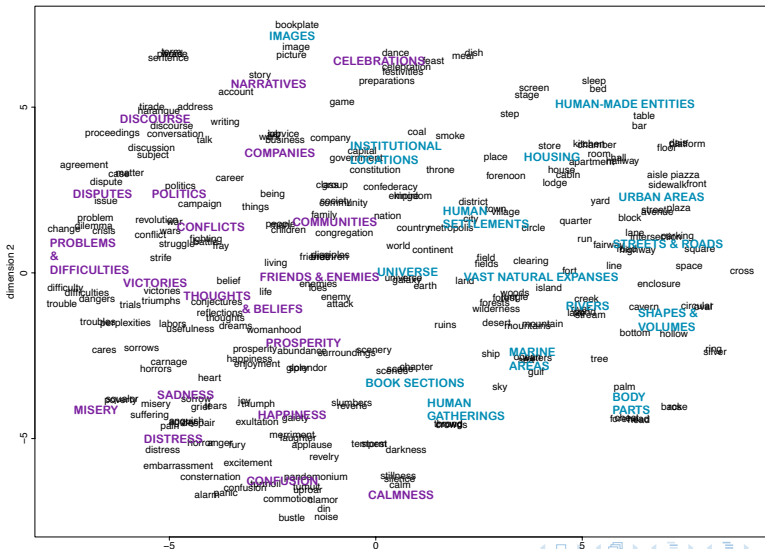
$$N_{center} = 4224$$

$$N_{heart} = 1347$$

building the reference SVS

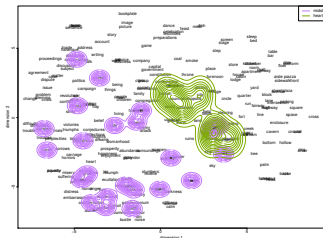
- I further trained word2vec on COHA
 - **hyperwords** (Levy et al 2015) → **SGNS**
 - **cds 0.75; win ±5; neg 15; dim 300** (Levy & Goldberg 2014)
 - all words whose frequency is less than 5 in the COHA were discarded
- the most distinctive landmark collocates of the 4 subschemas across the whole COHA were selected with log-likelihood ratio (DUNNING 1993) (LLR > 10)
- the word vectors of these distinctive collocates were retrieved and mapped onto a semantic vector space

the reference SVS

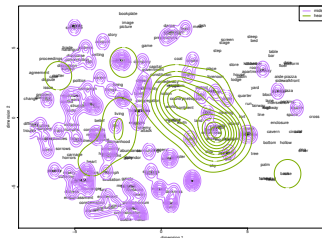


contour lines

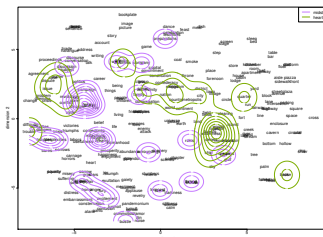
- contour lines are lines connecting data points of the same value.
- they are used to indicate how densely semantic areas have been populated for across four consecutive periods :
 - 1810s–1860s (period 1)
 - 1870s–1910s (period 2)
 - 1920s–1970s (period 3)
 - 1980s–2000s (period 4)
- the curves are indexed on period-wise LLR scores

MIDST_{cxn} and HEART_{cxn}

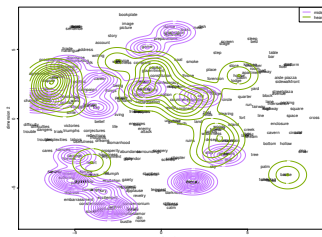
(a) Period 1



(b) Period 2



(c) Period 3



(d) Period 4

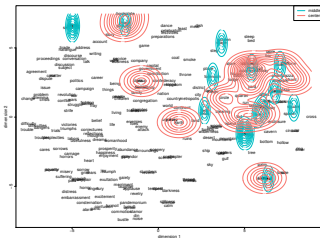
MIDST_{CXN} and HEART_{CXN}

A **complementary division of labour** is observed between MIDST_{CXN} and HEART_{CXN} : as HEART_{CXN} undergoes increased schematization, its contour lines hardly ever impinge upon those of MIDST_{CXN}.

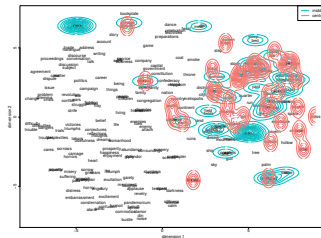
MIDST_{cxn} and HEART_{cxn}

- **MIDST_{cxn}** has shifted from the expression of space-related coincidence to two more schematic construals :
 - **the estranged construal** (the trajector appears as isolated in the landmark)
They were in the midst of the thickest forest.
[...] whom fate had surprised in the midst of other thought.
 - **the intensive construal** (the construction highlights a distinctive property of the landmark) *Then I had fallen in the midst of glory*
- **HEART_{cxn}** is schematizing towards **the intensive construal**
it is her dream to live in the heart of the country

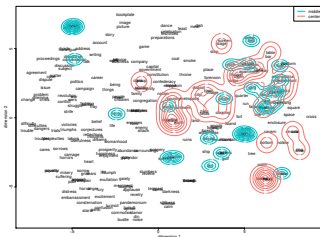
MIDDLE_{CXN} and CENTER_{CXN}



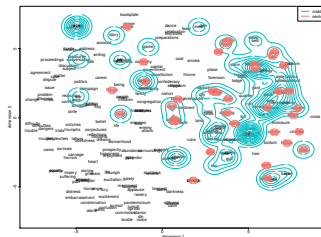
(a) Period 1



(b) Period 2



(c) Period 3



(d) Period 4

MIDDLE_{CXN} and CENTER_{CXN}

Compared to MIDST_{CXN} and HEART_{CXN}, the division of labor between MIDDLE_{CXN} and CENTER_{CXN} implies that the distribution of their respective landmarks is mostly **parallel**.

MIDDLE_{CXN} and CENTER_{CXN}

- **MIDDLE_{CXN}** has shifted from the expression of mere juxtaposition (the older role of MIDST_{CXN}) to the expression of **incongruous location**
- **CENTER_{CXN}** has shifted from the expression of central location to the expression of a **functional dependence** between the trajector and the landmark

Compare :

- (5) The car was parked **in the middle of the street**.
- (6) The car was parked **in the center of the street**.

and :

- (7) on that narrow avenue **in the heart of the city** is handled the enormous traffic between the twin cities
- (8) we cross the Gran Avenida, and continue along the wide Calle de Alma to the Puerta del Sol, **in the center of the city**.

Interim conclusion

- the same methods used to better understand the logic of vertical links can be used to better understand the horizontal relations between subshematic nodes in the construction
- the next step will consist in integrating the representation of nodes and links (vertical and horizontal) in a periodized network graph

word2vec or BERT ?

word2vec

- suffers from the 'one vector per word' issue (DESAGULIER 2019)
- an aggregation of all the contexts encountered by a single word form (including homonyms)
- in the case of polysemy, some linguists might be wary of the quality of the resulting vectors
- some linguists might be wary of the quality of the resulting vectors and turn to token-based distributional representations instead (FONTEYN 2021)

word2vec or BERT ?

BERT

- **In theory**, token-based DSMs such as BERT (DEVLIN et al. 2018) are optimal : token-based models do provide as many representations as there are contexts and do reflect semantic complexity.
- **In practice**, this comes at a huge computing cost. For this reason, most BERT models can only be trained on relatively small-sized corpora.

word2vec or BERT ?

Given the trade-off between computational cost and the semantic generalization requirement, **word2vec** is (still) a reasonable alternative :

- proper hyperparameter setup is well-document topic
- it is not as much of a 'black box' as state-of-the-art token-based DSMs
- the aggregated approach to meaning representation offered by type-based DSMs captures the average meaning of any given word type (ERK et PADÓ 2010, p. 92).

Conversely, most BERT distributions offer access to last two final hidden layers. When access is granted, there is still discussion as to what layer is relevant to the semanticist (MUN 2021).

Construction networks and graph theory

CxG adopts the assumptions and the terminology of graph theory :

- nodes, links, weights, centrality

Surprisingly, CxG supporters are extremely reluctant to embrace the maths and visualizations proposed by graph theory

- why?

thanks for your attention!

contact

`gdesagulier@univ-paris8.fr`

more details

`https://corpling.hypotheses.org/3583`

references I



BARONI, Marco, Georgiana DINU et Germán KRUSZEWSKI (2014). « Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. ». In : *ACL (1)*, p. 238-247.



BENGIO, Yoshua et al. (2003). « A neural probabilistic language model ». In : *Journal of Machine Learning Research* 3, p. 1137-1155.



COLLOBERT, Ronan et al. (2011). « Natural language processing (almost) from scratch ». In : *Journal of Machine Learning Research* 12, p. 2493-2537.



DESAGULIER, Guillaume (2019). « Can word vectors help corpus linguists ? » In : *Studia Neophilologica* 91.2, p. 219-240.



DEVLIN, Jacob et al. (2018). « Bert : Pre-training of deep bidirectional transformers for language understanding ». In : *arXiv preprint arXiv :1810.04805*.



DIESEL, Holger (2019). *The Grammar Network : How Linguistic Structure is Shaped by Language Use*. Cambridge : Cambridge University Press.

references II



DUNNING, Ted E (1993). « Accurate methods for the statistics of surprise and coincidence ». In : *Computational linguistics* 19.1, p. 61-74.



ERK, Katrin et Sebastian PADÓ (2010). « Exemplar-based models for word meaning in context ». In : *Proceedings of the acl 2010 conference short papers*, p. 92-97.



FIRTH, J. R. (1957). « A synopsis of linguistic theory 1930-55. ». In : *Studies in Linguistic Analysis (special volume of the Philological Society)*. T. 1952-59. Oxford : The Philological Society, p. 1-32.



FORTEYN, Lauren (2021). « Varying abstractions : a conceptual vs. distributional view on prepositional polysemy ». In : *Glossa : A journal of general linguistics* 6.1.



HAMILTON, William L, Jure LESKOVEC et Dan JURAFSKY (2016). « Diachronic word embeddings reveal statistical laws of semantic change ». In : *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, p. 1489-1501.



HARRIS, Zellig S. (1954). « Distributional structure ». In : *Word* 10.2-3, p. 146-162.

references III



HILPERT, Martin (2016). « Change in modal meanings ». In : *Constructions and Frames* 8.1, p. 66-85.



KULKARNI, Vivek et al. (2015). « Statistically significant detection of linguistic change ». In : *Proceedings of the 24th International Conference on World Wide Web*, p. 625-635.



MIKOLOV, Tomas et al. (2013). « Distributed representations of words and phrases and their compositionality ». In : *Advances in Neural Information Processing Systems*, p. 3111-3119.



MILLER, George A. et Walter G. CHARLES (1991). « Contextual correlates of semantic similarity ». In : *Language and Cognitive Processes* 6.1, p. 1-28.



MUN, Seongmin (2021). « Polysemy resolution with word embedding models and data visualization : the case of adverbial postpositions -ey, -eyse, and -(u)lo in Korean ». Thèse de doct. Paris Nanterre University : École Doctorale Connaissance, Langage et Modélisation.



PEREK, Florent (2016). « Using distributional semantics to study syntactic productivity in diachrony : A case study ». In : *Linguistics* 54.1, p. 149-188.

references IV



PEREK, Florent (2018). « Recent change in the productivity and schematicity of the way-construction : A distributional semantic analysis ». In : *Corpus Linguistics and Linguistic Theory* 14.1, p. 65-97.



SAGI, Eyal, Stefan KAUFMANN et Brady CLARK (2011). « Tracing semantic change with latent semantic analysis ». In : *Current Methods in Historical Semantics*, p. 161-183.



SMIRNOVA, Elena et Lotte SOMMERER (2020). « Introduction. The nature of the node and the network—Open questions in Diachronic Construction Grammar ». In : p. 2-42.



TRAUGOTT, Elizabeth Closs et Graeme TROUSDALE (2013). *Constructionalization and constructional changes*. T. 6. Oxford University Press.



ZEHENTNER, Eva et Elizabeth TRAUGOTT (2020). « Constructional networks and the development of benefactive ditransitives in English ». In : 27, p. 167-211.